

(19)



JAPANESE PATENT OFFICE

## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **07325583 A**(43) Date of publication of application: **12.12.95**

(51) Int. Cl

**G10H 7/08****G10H 7/00****G10L 3/02****G10L 9/14**(21) Application number: **05349245**(22) Date of filing: **28.12.93**(30) Priority: **14.04.93 US 93 48261**(71) Applicant: **YAMAHA CORP**(72) Inventor: **ZABIERU SERA  
KURISU UIRIAMUSU  
ROBAATO GUROSU  
AARINGU UORUDO**(54) **METHOD AND DEVICE FOR ANALYZING AND  
SYNTHESIZING SOUND**

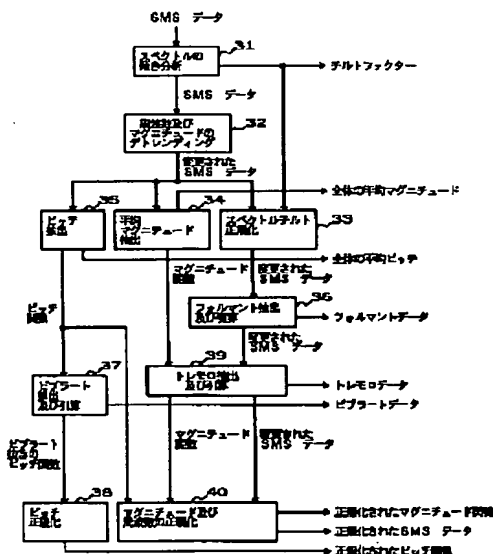
controlled characteristics.

COPYRIGHT: (C)1995,JPO

(57) Abstract:

**PURPOSE:** To adopt an analytic approach and to perform excellent sound control for controlling a musical characteristic of a synthesized sound in an analytic sound synthesis technique such as partial sound synthesis, etc., based on a spectrum modeling synthesis (SMS) technique.

**CONSTITUTION:** Characteristics (formant, vibrato, tremolo, spectrum, pitch change, etc.) related to prescribed elements are analyzed respectively from the analysis data showing plural components constituting an original sound waveform, and the data showing the analyzed characteristics are extracted as musical parameters. The characteristics corresponding to the extracted parameters are eliminated from the analysis data, and by the combination between the revised analysis data and the musical parameters, a sound waveform is expressed. The extracted musical parameters are variable controlled, and by adding them to the analysis data, the sound waveform is reproduced and synthesized based on the analysis data adding the



BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11) 特許番号

第2906970号

(45) 発行日 平成11年(1999) 6月21日

(24) 登録日 平成11年(1999) 4月2日

(51) Int.Cl.<sup>8</sup>

識別記号

F I

G 1 0 H 7/08  
7/00

G 1 0 H 7/00

5 3 2 Z

5 1 1 J

5 1 1 M

請求項の数12(全 39 頁)

(21) 出願番号 特願平5-349245

(22) 出願日 平成5年(1993)12月28日

(65) 公開番号 特開平7-325583

(43) 公開日 平成7年(1995)12月12日

審査請求日 平成8年(1996)2月7日

(31) 優先権主張番号 08/048, 261

(32) 優先日 1993年4月14日

(33) 優先権主張国 米国 (U S)

(73) 特許権者 000004075

ヤマハ株式会社

静岡県浜松市中沢町10番1号

(72) 発明者 ザビエル セラ

スペイン, バルセロナ 08025, ジョア

キム ルーイラ 10, 1イーアール 7  
エー

(72) 発明者 クリス ウィリアムス

アメリカ合衆国, カリフォルニア

94901, サンラファエル, ウィルキンス  
ストリート アパートメントハウス  
7

(74) 代理人 弁理士 飯塚 義仁

審査官 千葉 輝久

最終頁に続く

(54) 【発明の名称】 サウンドの分析及び合成方法並びに装置

1

(57) 【特許請求の範囲】

【請求項1】 オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、

前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第2のステップと、

抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除く第3のステップと、

前記時変動成分が取り除かれた分析データに基づきサウンド波形を合成するステップであって、該時変動成分が取り除かれた分析データ又は合成されたサウンド波形に

2

対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成する第4のステップとを備えたサウンドを分析し合成するための方法。

【請求項2】 前記第2のステップで抽出したサウンドパラメータを変更する第5のステップを更に備え、前記第4のステップでは、前記時変動成分が取り除かれた分析データ又は合成されたサウンド波形に対して、前記第5のステップで変更したサウンドパラメータに対応する特徴を付加するようにした請求項1に記載の方法。

【請求項3】 オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を

示すサウンドパラメータとして抽出する第2のステップと、  
抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除く第3のステップと

を備え、前記時変動成分が取り除かれた分析データと、前記サウンドパラメータとの組合せによって前記オリジナルサウンド波形を表現することを特徴とするサウンドを分析するための方法。

【請求項4】 オリジナルサウンドの分析に基づき該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する分析手段と、

前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータをサウンドパラメータとして抽出すると共に、抽出されたサウンドパラメータに基づき前記オリジナルサウンドの分析データから時変動成分を取り除くデータ処理手段と、

前記時変動成分が取り除かれた分析データと前記サウンドパラメータとを記憶する記憶手段と、

前記分析データとサウンドパラメータを前記記憶手段から読み出し、読み出した分析データとサウンドパラメータに基づきサウンド波形を合成する手段であって、該分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

【請求項5】 部分音を示すデータを含む波形分析データであって、オリジナルサウンドの分析データから時変動成分を取り除いてなるものと、オリジナルサウンドから抽出された所定のサウンド要素に関する特徴を示すサウンドパラメータとを記憶している記憶手段と、

前記波形分析データとサウンドパラメータを前記記憶手段から読み出す読み出し手段と、

読み出されたサウンドパラメータを変更するための制御を行なう制御手段と、

前記読み出した波形分析データを前記制御されたサウンドパラメータによって変更し、変更された波形分析データに基づきサウンド波形を合成する手段であって、該分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

【請求項6】 オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1の手段と、

前記第1の手段で提供される前記分析データに含まれる前記オリジナルサウンド波形のスペクトル分析データから該オリジナルサウンド波形のフォルマント構造を検出し、検出したフォルマントを記述するフォルマントパラメータを生成する第2の手段と、

前記スペクトル分析データから前記検出されたフォルマント構造を差引き、残余のスペクトルデータを生成する第3の手段と、

前記第1の手段で提供される前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第4の手段と、

抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除く第5の手段とを備え、前記残余のスペクトルデータと前記フォルマントパラメータと前記サウンドパラメータの組合せによって前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

【請求項7】 オリジナルサウンドの分析によって得た複数のサウンド部分を示す部分音データのセットを提供するものであり、各部分音データは周波数データを含み、前記部分音データのセットを時間関数で提供する第1の手段と、

20 前記部分音データにおける周波数データの時間関数からオリジナルサウンドにおけるビブラトを検出し、検出したビブラトを記述するパラメータを生成する第2の手段と、

前記部分音データにおける周波数データの時間関数から前記検出されたビブラトの特徴を取り除き、修正された周波数データの時間関数を生成する第3の手段とを備え、前記修正された周波数データの時間関数を含む前記部分音データと前記パラメータとの組合せによって時間的に変化する前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

30 【請求項8】 オリジナルサウンドの分析によって得た複数のサウンド部分を示す部分音データのセットを提供するものであり、各部分音データはマグニチュードデータを含み、前記部分音データのセットを時間関数で提供する第1の手段と、

前記部分音データにおけるマグニチュードデータの時間関数からオリジナルサウンドにおけるトレモロを検出し、検出したトレモロを記述するパラメータを生成する第2の手段と、

40 前記部分音データにおけるマグニチュードデータの時間関数から前記検出されたトレモロの特徴を取り除き、修正されたマグニチュードデータの時間関数を生成する第3の手段とを備え、前記修正されたマグニチュードデータの時間関数を含む前記部分音データと前記パラメータとの組合せによって時間的に変化する前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

【請求項9】 オリジナルサウンドのスペクトル構造を示すスペクトルデータを提供する第1の手段と、

50 前記スペクトルデータに基づき、そのスペクトルエンベ

ローブに概ね適合しているただ1本のチルトラインを検出し、検出したチルトラインの形状を示すチルトパラメータを生成する第2の手段と、

スペクトルの形状を制御するために、前記チルトパラメータを可変制御する第3の手段と、

制御されたチルトパラメータに基づき前記スペクトルデータのスペクトル構造を制御する第4の手段と、

制御されたスペクトルデータに基づきサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

【請求項10】 オリジナルサウンドを構成する部分音のスペクトルデータを複数の時間フレームに対応して提供する第1の手段と、

一連の時間フレームにおける前記部分音スペクトルデータにおける周波数データに基づき前記オリジナルサウンドの平均ピッチを検出し、ピッチデータを生成する第2の手段と、

前記ピッチデータを可変制御するための第3の手段と、制御されたピッチデータに応じて前記部分音スペクトルデータにおける周波数データを修正する第4の手段と、前記修正された周波数データを含む前記部分音スペクトルデータに基づき、可変制御されたピッチを持つサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

【請求項11】 オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、

前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、

前記データリストを参照して、任意の波形セグメントを選択するステップと、選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から抜き出すステップと、

抜き出したスペクトルデータ列を繰り返すことにより前記波形セグメントの繰り返しに対応するスペクトルデータ列を作成するステップと、

前記繰り返しに対応するスペクトルデータ列を使用して、延長された長さを持つサウンド波形を合成するステップとを備えたサウンドを分析し合成する方法。

【請求項12】 オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、

前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、

前記データリストを参照して、任意の波形セグメントを選択するステップと、

選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から取り去り、その前後で残された2つのスペクトルデータ列を接続し、短縮されたスペクトルデータ列を作成するステップと、

前記短縮されたスペクトルデータ列を使用して、短縮された長さを持つサウンド波形を合成するステップとを備えたサウンドを分析し合成する方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、サウンド特に楽音あるいは人声音などのような音楽サウンドの分析及び合成方法並びに装置に関し、更には、スペクトル・モデリング・合成 (Spectral Modeling Synthesis) 技術を用いた音楽シンセサイザにおける様々な改良に関する。

【0002】

【従来の技術】 スペクトル・モデリング・合成 (以下、SMSと略称する) 技術を用いた音楽シンセサイザの従来技術は、本願の発明者の一人であるザビエル・セラ (Xavier Serra) の執筆に関わる「確定的成分とストカスティック成分の分解に基づくサウンドの分析/変換/合成のためのシステム」 ("A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition") と題する1989年10月発表のスタンフォード大学博士論文に示されている。また、同じくザビエル・セラの発明に関わる「確定的波形とストカスティック波形の組合せによる音楽シンセサイザ」 ("Musical Synthesizer Combining Deterministic and Stochastic Waveforms") と題する米国特許第5,029,509号にも示されており、また、上記米国特許に対応する国際出願公開番号W090/13887にも開示されている。

【0003】 SMS技術は、サウンドが2つのタイプの成分、すなわち確定的成分 (a deterministic component) とストカスティック成分 (a stochastic component; 確率的な若しくは不規則的な成分)、で構成され则认为モデルを使用する楽音の分析及び合成技術である。確定的成分は、一連のシナソイド (sinusoid; 正弦波の形状で変化する波形) で表わされ、各シナソイド毎に振幅と周波数関数を持つ。つまり、確定された振幅と周波数を持つスペクトル成分である。ストカスティック成分は、マグニチュード・スペクトルエンベロープで表わされる。例えば、オリジナル波形のスペクトルから確定的成分のスペクトルを差し引いた結果であるところの残差スペクトルを、スペクトルエンベロープで表現したものがストカスティック成分である。サウンドの分析と合成は、一連の時間フレームにおける各時間フレーム毎に夫々行なわれる。

【0004】各時間フレーム毎の分析データは、夫々が特定の周波数と振幅値を持つ1揃いの部分音 (partial; パーシャル) と、周波数領域のエンベロープとによって、下記数1のように、表現される。

【0005】

【数1】

$$\begin{aligned} a_n(t), f_n(t) & \quad \text{ただし、} n=0, \dots, N-1 \\ e_m(t) & \quad \text{ただし、} m=0, \dots, M-1 \end{aligned}$$

【0006】ここで、 $f$  は、特定のフレームを示す。 $a_n(t)$  と  $f_n(t)$  は、フレーム  $t$  における各部分音の振幅及び周波数を示し、確定的成分に対応する。 $N$  は、そのフレームにおける部分音の数である。 $e_m(t)$  は、ストカスティック成分に対応するスペクトルエンベロープであり、 $m$  はブレイクポイント番号、 $M$  はそのフレームにおけるブレイクポイント数である。

【0007】

【発明が解決しようとする課題】このようなSMS技術に基づく楽音合成は、圧縮された分析データを使用して極めて高品質のサウンド波形を合成することができるという利点を持つ。また、サウンド合成に使用する分析データを、ユーザーが自由に制御することにより、幅広い多様なニューサウンドを作り出すことができる可能性を秘めているものである。そこで、SMS技術に基づく楽音合成技術において、様々な音楽的制御のための具体的手法を確立することが望まれていた。

【0008】一方、オリジナルサウンド波形をフーリエ変換その他の技術によって分析して部分音スペクトルデータを得て、これをメモリに記憶し、メモリから読み出した部分音スペクトルデータを逆フーリエ変換することによりサウンド波形を合成する技術それ自体もよく知られている。しかし、従来知られた部分音合成技術は、単なる合成技術にすぎず、合成しようとするサウンドの音楽的特徴を制御するために分析的なアプローチを採用するものではなかった。

【0009】音楽シンセサイザにおける一つの技術的課題として、人声音を如何にして合成するかというものがある。従来知られたボーカル音合成技術の多くは、ボーカル・モデルに基づくものである。すなわち、振動信号を時変動するフィルタに通すものである。このモデルは、高品質のサウンドを生成することができず、また、融通性に欠けているものである。また、従来のボーカル音合成技術の大多数は、分析に基づくものではなく、単なる合成技術である。すなわち、或る与えられたシンガー（歌手）に基づいてモデル形成できるものではない。また、従来の技術では、記録したシンガー音からビブラートを取り除くための方法が提案されていない。

【0010】この発明は上述の点に鑑みてなされたものであり、その1つの目的は、SMS技術に基づく楽音合成技術あるいは部分音合成技術又はその他の分析的なサ

ウンド合成技術において、合成しようとするサウンドの音楽的特徴を制御するために分析的なアプローチを採用することにより、良好なサウンド制御を達成しうることによることである。また、この発明の目的は、SMS技術を基にしたサウンドの合成及び分析において、様々な改良を提案し、その実用性を高めることにある。

【0011】更に、この発明の他の目的は、オリジナルサウンド波形の分析データからフォルマントの特徴を抽出し、制御し、サウンド波形の合成のために利用するための技術を提供することにある。更に、この発明の他の目的は、オリジナルサウンド波形の分析データからビブラート又はトレモロの特徴を抽出し、制御し、サウンド波形の合成のために利用するための技術を提供することにある。更に、この発明の他の目的は、オリジナルサウンド波形の分析データからスペクトルチルトの特徴を抽出し、制御し、サウンド波形の合成のために利用するための技術を提供することにある。

【0012】更に、この発明の他の目的は、オリジナルサウンド波形の分析データからピッチを抽出し、制御し、可変ピッチ制御したサウンド波形を合成するために利用するための技術を提供することにある。更に、この発明の他の目的は、オリジナルサウンド波形の分析データからビブラートのような低周波域の変動を検出することにより特定の波形セグメントを抽出し、抽出した波形セグメントを制御し、発音時間長を延長又は短縮したサウンド波形を合成するために利用するための技術を提供することにある。更に、この発明の他の目的は、SMS技術とデジタルウェーブガイド技術とを融合した新規なサウンド合成技術を提供することにある。更に、この発明の他の目的は、SMS技術を使用した分析的な手法により、高品質のボーカルフレーズ音声を合成することを提案することにある。

【0013】

【課題を解決するための手段】 上述の目的を達成するために、第1の観点に従えば、この発明に係るサウンドを分析し合成するための方法は、オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第2のステップと、抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除く第3のステップと、前記時変動成分が取り除かれた分析データに基づきサウンド波形を合成するステップであって、該時変動成分が取り除かれた分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成する第4のステップとを備える。

【0014】 上述の目的を達成するために、第2の観点に従えば、この発明に係るサウンドを分析し合成するための方法は、前記第2のステップで抽出したサウンドパラメータを変更する第5のステップを更に備え、前記第4のステップでは、前記時変動成分が取り除かれた分析データ又は合成されたサウンド波形に対して、前記第5のステップで変更したサウンドパラメータに対応する特徴を付加するようにしたことを特徴とする。

【0015】 上述の目的を達成するために、第3の観点に従えば、この発明に係るサウンドを分析するための方法は、前記第1のステップと、前記第2のステップと、前記第3のステップとを備え、前記時変動成分が取り除かれた分析データと、前記サウンドパラメータとの組合せによって前記オリジナルサウンド波形を表現することを特徴とする。

【0016】 上述の目的を達成するために、第4の観点に従えば、この発明に係るサウンド波形合成装置は、オリジナルサウンドの分析に基づき該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する分析手段と、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータをサウンドパラメータとして抽出すると共に、抽出されたサウンドパラメータに基づき前記オリジナルサウンドの分析データから時変動成分を取り除くデータ処理手段と、前記時変動成分が取り除かれた分析データと前記サウンドパラメータとを記憶する記憶手段と、前記分析データとサウンドパラメータを前記記憶手段から読み出し、読み出した分析データとサウンドパラメータに基づきサウンド波形を合成する手段であって、該分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するサウンド合成手段とを備えたことを特徴とする。

【0017】 上述の目的を達成するために、第5の観点に従えば、この発明に係るサウンド波形合成装置は、部分音を示すデータを含む波形分析データであって、オリジナルサウンドの分析データから時変動成分を取り除いてなるものと、オリジナルサウンドから抽出された所定のサウンド要素に関する特徴を示すサウンドパラメータとを記憶している記憶手段と、前記波形分析データとサウンドパラメータを前記記憶手段から読み出す読み出し手段と、読み出されたサウンドパラメータを変更するための制御を行なう制御手段と、前記読み出した波形分析データを前記制御されたサウンドパラメータによって変更し、変更された波形分析データに基づきサウンド波形を合成する手段であって、該分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するサウンド合成手段とを備える。

【0018】 上述の目的を達成するために、第6の観

点に従えば、この発明に係るサウンド波形合成装置は、オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1の手段と、前記第1の手段で提供される前記分析データに含まれる前記オリジナルサウンド波形のスペクトル分析データから該オリジナルサウンド波形のフォルマント構造を検出し、検出したフォルマントを記述するフォルマントパラメータを生成する第2の手段と、前記スペクトル分析データから前記検出されたフォルマント構造を差引き、残余のスペクトルデータを生成する第3の手段と、前記第1の手段で提供される前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第4の手段と、抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除く第5の手段とを備え、前記残余のスペクトルデータと前記フォルマントパラメータと前記サウンドパラメータの組合せによって前記オリジナルサウンド波形を表現することを特徴とする。このサウンド波形合成装置は、更に、前記フォルマントを制御するために前記フォルマントパラメータを可変制御する第6の手段と、前記フォルマントパラメータに基づきフォルマント構造を再生し、再生されたフォルマント構造を前記残余のスペクトルデータに付加し、制御されたフォルマント構造を有するスペクトルデータを作成する第7の手段と、前記第7の手段で作成されたスペクトルデータに基づきサウンド波形を合成するサウンド合成手段とを具備してよい。

【0019】 上述の目的を達成するために、第7の観点に従えば、この発明に係るサウンド波形合成装置は、オリジナルサウンドの分析によって得た複数のサウンド部分を示す部分音データのセットを提供するものであり、各部分音データは周波数データを含み、前記部分音データのセットを時間関数で提供する第1の手段と、前記部分音データにおける周波数データの時間関数からオリジナルサウンドにおけるビブラートを検出し、検出したビブラートを記述するパラメータを生成する第2の手段と、前記部分音データにおける周波数データの時間関数から前記検出されたビブラートの特徴を取り除き、修正された周波数データの時間関数を生成する第3の手段とを備え、前記修正された周波数データの時間関数を含む前記部分音データと前記パラメータとの組合せによって時間的に変化する前記オリジナルサウンド波形を表現することを特徴とする。このサウンド波形合成装置は、更に、ビブラートを制御するために前記パラメータを可変制御する第4の手段と、前記パラメータに基づきビブラート関数を発生し、発生されたビブラート関数によって前記修正された周波数データの時間関数にビブラートを付与する第5の手段と、ビブラート付与された周波数デ

ータの時間関数を含む前記部分音データに基づきサウンド波形を合成するサウンド合成手段とを備えていてよい。

【0020】上述の目的を達成するために、第8の観点に従えば、このサウンド波形合成装置において、前記部分音データにおけるマグニチュードデータの時間関数からオリジナルサウンドにおけるトレモロを検出し、これにより上記ビブラートの場合と同様の処理をしてもよい。そうすると、トレモロの抽出と、その可変制御、及びそれに基づくサウンド波形合成が可能である。

【0021】上述の目的を達成するために、第9の観点に従えば、この発明に係るサウンド波形合成装置は、オリジナルサウンドのスペクトル構造を示すスペクトルデータを提供する第1の手段と、前記スペクトルデータに基づき、そのスペクトルエンベロープに概ね適合しているただ1本のチルトラインを検出し、検出したチルトラインの形状を示すチルトパラメータを生成する第2の手段と、スペクトルの形状を制御するために、前記チルトパラメータを可変制御する第3の手段と、制御されたチルトパラメータに基づき前記スペクトルデータのスペクトル構造を制御する第4の手段と、制御されたスペクトルデータに基づきサウンド波形を合成するサウンド合成手段とを備え。

【0022】上述の目的を達成するために、第10の観点に従えば、この発明に係るサウンド波形合成装置は、オリジナルサウンドを構成する部分音のスペクトルデータを複数の時間フレームに対応して提供する第1の手段と、一連の時間フレームにおける前記部分音スペクトルデータにおける周波数データに基づき前記オリジナルサウンドの平均ピッチを検出し、ピッチデータを生成する第2の手段と、前記ピッチデータを可変制御するための第3の手段と、制御されたピッチデータに応じて前記部分音スペクトルデータにおける周波数データを修正する第4の手段と、前記修正された周波数データを含む前記部分音スペクトルデータに基づき、可変制御されたピッチを持つサウンド波形を合成するサウンド合成手段とを備える。

【0023】上述の目的を達成するために、第11の観点に従えば、この発明に係るサウンドを分析し合成する方法は、オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、前記データリストを参照して、任意の波形セグメントを選択するステップと、選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から抜き出すステップと、抜き出したスペクトルデータ列を繰

り返すことにより前記波形セグメントの繰返しに対応するスペクトルデータ列を作成するステップと、前記繰返しに対応するスペクトルデータ列を使用して、延長された長さを持つサウンド波形を合成するステップとを備える。上記方法において、更に、前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナル波形から引いた残りである残差成分波形に対応するストカスティックデータを、複数の時間フレームに対応してシリーズで提供するステップと、前記選択した波形セグメントに対応する前記ストカスティックデータシリーズを前記オリジナル波形のストカスティックデータシリーズから抜き出すステップと、抜き出したストカスティックデータシリーズを繰返しことにより前記波形セグメントの繰返しに対応するストカスティックデータシリーズを作成するステップと、前記繰返しに対応するストカスティックデータシリーズを使用して、延長された長さを持つストカスティック波形を合成し、これを前記サウンド波形に組み込むステップとを備えていてもよい。

【0024】上述の目的を達成するために、第12の観点に従えば、この発明に係るサウンドを分析し合成する方法は、オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、前記データリストを参照して、任意の波形セグメントを選択するステップと、選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から取り去り、その前後で残された2つのスペクトルデータ列を接続し、短縮されたスペクトルデータ列を作成するステップと、前記短縮されたスペクトルデータ列を使用して、短縮された長さを持つサウンド波形を合成するステップとを備える。上記方法において、更に、前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナル波形から引いた残りである残差成分波形に対応するストカスティックデータを、複数の時間フレームに対応して順次に提供するステップと、前記選択した波形セグメントに対応する前記ストカスティックデータ列を前記オリジナル波形のストカスティックデータ列から取り去り、その前後で残された2つのストカスティックデータ列を接続し、短縮されたストカスティックデータ列を作成するステップと、前記短縮されたストカスティックデータ列を使用して短縮された長さを持つストカスティック波形を合成し、これを前記サウンド波形に組み込むステップとを更に備えていてよい。

【0025】

【作用】 前記第1の観点乃至第5の観点のいずれかに従う方法または装置によれば、オリジナルサウンドの分

析データから所定のサウンド要素に関する特徴を分析し、分析した該特徴を示すデータをサウンドパラメータとして抽出するようにしたので、例えばフォルマントやビブラートなどのような様々なサウンド要素に関して、オリジナルの特徴を示している品質のよいサウンドパラメータを得ることができる。従って、このパラメータをサウンド波形合成に際して利用すれば、品質のよい各種音楽的特徴の合成を行なうことができる。しかも、サウンドパラメータとして分析データから分離抽出されているため、その可変制御が容易であり、ユーザーによる自由な音楽制御に適したものである。また、抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除くようにしているため、分析データの構造が簡単化され、データ圧縮が期待できるものである。また、時変動成分が取り除かれた分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するようにしているので、オリジナルの特徴を示している品質のよいサウンド波形に対して任意の時間変化特性を付与することができ、品質と制御性に優れたサウンド波形合成を行なうことができる。このように、サウンドパラメータを分析データから抽出分離し、時変動成分が取り除かれた分析データとサウンドパラメータの組み合わせによってオリジナルサウンド波形を表現するデータを提供し、これに基づきサウンド波形を合成する技術は、種々の効果が期待できるものである。前記第6の観点乃至第12の観点のいずれかに従う装置または方法によれば、各種のサウンドパラメータ（フォルマント、ビブラート、トレモロ、スペクトルなど）の抽出とそれに基づく波形合成・制御が達成される。

【0026】

【実施例】以下、この発明の実施例を添付図面を参照して詳細に説明しよう。

【全体説明】図1は、この発明の一実施例に係る音楽シンセサイザーの全体図である。このシンセサイザーは、大別して、オリジナルサウンドの分析を行なう分析部10と、分析された表現物すなわち分析データからサウンドを合成する合成部11とを含んでいる。オリジナルサウンドはマイクロフォン12によって外部からピックアップして、分析部10に入力するようにしてよいし、その他の適宜の方法で分析部10に導入してもよい。このシンセサイザーにおける分析と合成の両方が、前述の米国特許第5,029,509号にその基本原理が示されたようなSMS（スペクトル・モデリング・合成）技術を基にしているものである。なお、分析されたデータがすでにシンセサイザーのメモリ内にストアされていてもよく、その場合は分析部10はオプション的であってよい。このシンセサイザーは、シンギング・ボイス（人の歌声）若しくはボーカル・フレーズ（人声音フレーズ）

の分析及び合成に適しているシンギング・シンセサイザーとして構成してよい。しかし、本発明は、シンギングボイスに限らず、自然楽器音やその他の楽音／サウンド一般の分析と合成に応用可能である。

【0027】以下で説明する実施例においては、SMS分析に関して或るいくつかの改良がなされている。そのような改良は、シンギング・ボイス若しくはボーカル・フレーズの分析と合成に適しているものであるが、サウンド一般の分析と合成にも適するものである。そのような改良の1つとして、SMS分析データから所定のサウンド要素に関する特徴を分析し、分析した特徴を示すデータをサウンドパラメータとして抽出するための処理が分析部10で行なわれる。このサウンドパラメータを以下では音楽パラメータという。抽出された音楽パラメータは、合成部11に与えられ、サウンド合成の際にユーザーによって操作することができるようになっている。すなわち、ユーザーは、合成しようとするサウンドを好みに応じて変更制御しようとする場合、特殊なSMS分析データのフォームからなるパラメータで相互作用する必要がなく、なじみの深い従前の音楽制御情報に対応するフォームからなる音楽パラメータで相互作用すればよいことになり、便利である。そのような音楽パラメータとは、例えば、トーンピッチ、ビブラート、トレモロ、などのような音楽要素又は楽音要素に対応するパラメータである。そのために、相互作用的な編集機器群13や音楽コントローラ群14を装備してよい。

【0028】編集機器群13は、各種のコンピュータ端末機器（入力キーボードや、ディスプレイ、マウスその他）であってよい。音楽コントローラ群14は、音階音を指定するためのキーボードや、音色を選択若しくは設定するためのパネルスイッチ群や、各種の楽音効果を選択制御するためのスイッチ群や、ユーザーの自由な意志に従って楽音制御を行なうための各種の操作子群などを含んでいてよい。この音楽コントローラ群14の中には、ユーザーのボイス（人の音声）で楽音を制御するものを含んでいてもよいし、ボディアクションやプレスで制御するものを含んでいてもよい。ユーザーにより操作可能なこれらの編集機器群13及びコントローラ群14と、合成部11との間には、音楽パラメータインターフェース部15が設けられ、パラメータのやり取りや情報の翻訳が適切に行なわれる。

【0029】以下、図2以降の図を参照して、このシンセサイザーの詳細例を更に詳しく説明するが、提示した各部の詳細図は機能ブロック図である。図示された各機能を実現する手段を、ディスクリート回路によって構成してもよいし、マイクロコンピュータを使用したソフトウェア処理によって構成してもよい。また、このシンセサイザーは、以下で説明する改良に関わる全ての機能を併せ持っている必要はなく、任意の1つの機能のみ持っているだけでもよい。



【0030】〔分析部の説明〕図2は、分析部10の一例を示すブロック図である。オリジナルサウンド信号が入力されるSMS分析器20は、前述の米国特許第5,029,509号に示されたようなSMS分析技術に従ってオリジナルサウンドのSMS分析を行なうものである。このSMS分析器20の具体的構成としては、例えばその米国特許の第1図に示されたような構成を参照することができる。しかし、便宜のために、SMS分析器20の基本構成例をブロック20内に概略的に示す。

#### 【0031】－SMS分析器－

SMS分析器20において、入力サウンド信号は、最初に時間窓処理部20aで処理される。ここでは、入力サウンド信号を、時間窓と言われる一連の時間フレームに分ける処理が行なわれる。次の周波数分析部20bでは、各時間フレーム毎のサウンド信号を分析し、1組のマグニチュードスペクトルデータを発生する。例えば、高速フーリエ変換器（FFT）による分析により、複素数スペクトルを発生し、複素数-実数変換器でこれをマグニチュードスペクトルに変換するようにしてもよいし、その他の周波数分析方法を使用してもよい。

【0032】線スペクトル抽出部20cでは、分析されたオリジナルサウンドの1組のマグニチュードスペクトルから部分音の線スペクトルを抽出する。例えば、分析されたオリジナルサウンドの1組のマグニチュードスペクトルにおけるピークを検出し、これらのピークに対応する特定の周波数と振幅値すなわちマグニチュード値を持つスペクトルを線スペクトルとして抽出する。これらの抽出された線スペクトルが確定的成分に対応する。抽出された1つの線スペクトルすなわち確定的成分は、特定の周波数を示すデータと、その振幅値すなわちマグニチュード値を示すデータのペアからなっていてよいし、更にはそのデータペアに位相を示すデータが加わっていてもよい。これらの部分音の線スペクトルデータは、各時間フレームに対応して時系列的に得られるものであり、そのような時系列的な線スペクトルデータセットを夫々周波数トラジェクトリ（trajectory；軌跡若しくは遍歴）、マグニチュードトラジェクトリ、位相トラジェクトリと呼んでいる。

【0033】残差スペクトル生成演算部20dでは、各時間フレーム毎に、オリジナルサウンドの1組のマグニチュードスペクトルから、前記抽出された線スペクトルを引算し、残差スペクトルを生成する。この場合、前記米国特許に示されているように、前記抽出された線スペクトルに基づき確定的成分の波形を合成し、これを再分析して線スペクトルの再抽出を行ない、再抽出した線スペクトルをオリジナルサウンドの1組のマグニチュードスペクトルから引算する処理を行なってもよい。

【0034】次の残差スペクトルエンベロープ発生器20eでは、各時間フレーム毎に、残差スペクトルをエンベロープによって表現する処理を行なう。この残差スペ

クトルエンベロープは、例えば、線セグメント近似物の形でデータ表現することができるので、データ圧縮の促進に寄与する。一連の時間フレームに対応して発生される残差スペクトルエンベロープは、ストカスティック成分に対応するものである。SMS分析器20で得られる、確定的成分に対応する周波数トラジェクトリ及びマグニチュードトラジェクトリ（更に位相トラジェクトリを含んでいてもよい）と、ストカスティック成分に対応する残差スペクトルエンベロープとを総称して、以下ではSMSデータと呼ぶ。

#### 【0035】－SMSデータ処理の概略－

SMSデータ処理部30では、SMS分析器20で得られたSMSデータに対して適宜の処理を施す。ここでの処理は、大別して2種類ある。1つは、SMSデータを適宜に処理することにより、変更されたされたSMSデータを得ること。もう1つは、SMSデータから各種の音楽パラメータを抽出することである。データ処理ブロック30aでは、確定的成分に対応する周波数トラジェクトリ及びマグニチュードトラジェクトリ（更に位相トラジェクトリを含んでいてもよい）について上述のデータ処理を行なう。データ処理ブロック30bでは、ストカスティック成分に対応する残差スペクトルエンベロープについて上述のデータ処理を行なう。

【0036】SMSデータ処理部30における処理によって得られた、処理済みの又は変更されたSMSデータと、各種の音楽パラメータは、データメモリ100において各フレームに対応してストアされる。SMSデータ処理部30において行なう処理は、色々あるが、発明の実施にあたってはそのすべてを行なう必要はなく、適宜選択して実施してよい。処理が施されなかったSMSデータに関しては、分析器20から与えられたものと同じものがデータメモリ100にストアされるであろう。

【0037】SMSデータ処理部30において行なわれる各種処理の概略について図3を参照して説明する。ただし、図3は、SMSデータ処理部30において行なわれるすべての処理を紹介するものではなく、いくつかの代表的な処理について示している。前述のように、図3に示されたすべての処理を実施する必要はなく、実施にあたって不要なステップは適宜省略してよい。図3に示された処理のいくつかは追って更に詳しく説明される。また、図3に示されたなかった処理についても、追って詳しく説明されるものがある。

【0038】ステップ31：スペクトルの傾き分析  
この処理の基本思想は、マグニチュードとスペクトルのチルトすなわち傾きとの相関を見つけたことである。ここで、チルトとはスペクトルの全体的なスロープのことである。すなわち、チルトとは、各ハーモニックピークの頂部を概ね結んだ一直線状のスロープである。典型的には、音楽サウンドにおいて、チルトがより小さい場合、より高いハーモニックの振幅が相対的に高めら

れ、その結果、より明るい感じのサウンドをもたらす。このスペクトルの傾き分析処理では、“チルトファクター”と称する単一の数値データを求める。このチルトファクターは、マグニチュードとスペクトルのチルトとの間の相関を表わしている。このチルトファクターは、各フレーム毎に求められる。各フレーム毎に求めたチルトファクターを使用して、どのフレームに対しても共通の単一のチルトファクターを求めるための“スペクトルチルト正規化”が後のステップで行なわれる。チルトファクターは音楽パラメータの一種といつてよいものである。これによって、1つのチルトファクターをユーザーが自由に制御することで、SMSにより合成されるサウンドの特性を、ユーザーの意志を的確に反映してかつ自由に制御できるものとなる。

【0039】ステップ32：周波数及びマグニチュードのデトレンドリング (de-trending; 癖取り除き)  
記録したオリジナルサウンドは、その安定状態において、クレッシェンドやデクレッシェンドのような音量変化、又はわずかなピッチ変化を持っているのが普通である。ところで、記録した波形データの持続時間よりも長い時間だけサウンドを再生発音することを可能にする技術として、ループ処理といわれる繰返し発音処理を安定状態において行なうことが知られている。そのようなループ処理にあたって、ループする波形データ区間において音量やピッチの変動があると、ループポイント(繰返しをつなぎポイント)で目立った不連続が生じたり、ループによる不自然な周期性が目立ったりするので好ましくない。そこで、この問題を解決するために、このデトレンドリング処理では、SMSデータにおけるその種の変動を取り除き、サウンドの安定状態での全体的な傾向(トレンド)を可能な限り平坦にするよう処理する。ただし、ビブラートやサウンドの微変動は取り除かずに残しておく。

【0040】ステップ33：スペクトルチルト正規化  
ここでは、各フレーム毎に求めたチルトファクターを使用して、どのフレームに対しても共通の単一のチルトファクターを求める。これにより、ユーザーによる制御対象であるチルトファクターは、時間フレームに関係なく単一となるので、制御性が向上する。

【0041】ステップ34：平均マグニチュード抽出  
ここでは、各フレーム毎に、全ての確定的信号のマグニチュード値の平均値を計算する。すなわち、1つのフレームについては、全ての部分音成分のマグニチュード値を加算し、その加算値を部分音成分の数で割る。こうして得た各フレーム毎の平均マグニチュードをマグニチュード関数と呼ぶ。このマグニチュード関数は、確定的成分によって代表されるサウンドの音量の時変動を示している。さらには、これらのフレーム毎の平均マグニチュードから、全体の平均マグニチュードを計算する。全体の平均マグニチュードは、サウンドの安定状態につい

て計算される。この全体の平均マグニチュードは、安定状態における該サウンドの代表的音量レベルを示している。

#### 【0042】ステップ35：ピッチ抽出

ここでは、各フレーム毎のピッチが計算される。これは、1つのフレームについては、SMSデータにおける最初のいくつかの、つまり低次の、部分音成分を使用して、重み付けされた平均ピッチを計算することにより行なう。この重みづけにあたっては、重み付けファクターとして、各部分音成分のマグニチュード値を使用する。こうして求めた平均ピッチが、そのフレームにおけるサウンドのピッチと呼ばれる。こうして得た各フレーム毎の平均ピッチをピッチ関数と呼ぶ。このピッチ関数は、確定的成分によって代表されるサウンドのピッチの時変動を示している。さらには、これらのフレーム毎の平均ピッチから、全体の平均ピッチを計算する。全体の平均ピッチは、サウンドの安定状態について計算される。この全体の平均ピッチは、安定状態における該サウンドの代表的ピッチを示している。

#### 20 ステップ36：フォルマント抽出及び引算

この基本思想は、SMSデータからフォルマントを抽出し、抽出したフォルマントをSMSデータから引算することである。その結果得られる変更されたSMSデータにおける全ての部分音成分が似たようなマグニチュード値を持つことになる。つまり、スペクトル形状が平坦になる。抽出したフォルマントを表現するフォルマントデータは、後段の合成段階で利用される。このフォルマントデータは、音楽パラメータの一種といつてよいものである。これによって、フォルマントデータをユーザーが自由に制御することで、SMSにより合成されるサウンドの特性を、ユーザーの意志を的確に反映してかつ自由に制御できるものとなる。

#### 30 【0043】ステップ37：ビブラート抽出及び引算

ここでは、上記ステップ35で求めたピッチ関数から、ビブラートのかかっている部分を抽出し、抽出したビブラート成分をピッチ関数から引算する。抽出したビブラートを表現するビブラートデータは、後段の合成段階で利用される。ビブラートデータも、音楽パラメータの一種といつてよく、ユーザーによるビブラートの容易な制御を可能にする。

#### 40 【0044】ステップ38：ピッチ正規化

ここでは、上記ステップ37から出力されるビブラート抜ききのピッチ関数における各フレームの平均ピッチから前記全体平均ピッチを引き算することにより、正規化されたピッチ関数を得る。

#### 50 【0045】ステップ39：トレモロ抽出及び引算

ここでは、上記ステップ34で求めたマグニチュード関数から、トレモロのかかっている部分を抽出し、抽出したトレモロ成分をマグニチュード関数から引算する。こうして、トレモロデータとトレモロ成分を除去したマグ

ニチュード関数とを得る。また、SMSデータにおけるマグニチュードトラジェクトリからもトレモロ成分を除去し、かつ、ストカスティックゲイン（各フレーム毎の残差スペクトルエンベロープのゲイン）からトレモロ成分を除去してもよい。トレモロデータも、音楽パラメータの一種といってよく、ユーザーによるトレモロの容易な制御を可能にする。

【0046】ステップ40：マグニチュード及び周波数の正規化

ここでは、SMSデータを正規化する処理を行なう。周波数データは、ステップ35で抽出されたピッチ関数によって、各部分音成分毎の周波数トラジェクトリを、その部分音成分分だけ、割算することによって正規化される。これにより、各部分音成分の演算結果は、1に近い周波数値を持つようになる。マグニチュードデータは、マグニチュードトラジェクトリから、上記マグニチュード関数を引算することによって正規化する。ストカスティックデータについては、安定状態におけるストカスティックゲイン（各フレーム毎の残差スペクトルエンベロープのゲイン）の平均値を求め、これを基準値として、各フレーム毎の残差スペクトルエンベロープのゲインから引算することにより正規化してよい。こうして、正規化されたSMSデータを得るようにしてよい。また、マグニチュード関数に関しても、全体平均マグニチュードを基準にして正規化し、正規化されたマグニチュード関数を得るようにしてよい。

【0047】上述したようなSMSデータ処理部30における各処理によって得られた、処理済みの、すなわち変更された又は正規化されたSMSデータと、各種の音楽パラメータは、前述したように、データメモリ100において各フレームに対応してストアされる。前述したように、本発明の実施にあたっては、上述した各処理はオプションであるため、例えば上記ステップ40のような正規化処理を行なった場合は正規化されたSMSデータがデータメモリ100にストアされるが、行なわなかった場合は、単に変更されたSMSデータがデータメモリ100にストアされる。また、変更も正規化も行なわなかった場合は、SMS分析器20で分析されたままのSMSデータがデータメモリ100にストアされるであろう。

【0048】〔合成部の説明〕図4は、合成部11の一例を示すブロック図である。データメモリ100は、図2に示されたものと同じものであり、上記のように、各フレームについての処理済みのSMSデータと抽出された各種の音楽パラメータがストアされている。これらのデータは、1つのオリジナルサウンドに対応するものだけに限らず、多数の異なるオリジナルサウンドに対応するものをストアするようにしてもよいのは勿論である。

【0049】再生処理部50は、所望のサウンドを再生するために、データメモリ100からストアされたデー

タの読み出しを行なう処理と、読み出したSMSデータと音楽パラメータに基づく、追って述べるような様々なデータ操作処理を行なう。また、図1に示された編集機器群13や音楽コントローラ群14によって発生された制御パラメータを含む各種の音楽パラメータがこの再生処理部50に与えられ、この再生処理部50における各種処理をユーザーの制御に従って行なえるようにしている。例えば、ユーザーによって、望みのボイス若しくは音色を選択すると、このボイス若しくは音色に対応する1つのオリジナルサウンドに対応する一揃いのデータを、データメモリ100から読み出し可能にする。それから、ユーザーによって、発音開始指示が与えられると、時間フレームのシーケンスがスタートし、上記読み出し可能にされた一揃いのデータのうち、該シーケンスによって指定される特定のフレームについてのSMSデータと各種パラメータがデータメモリ100から読み出される。こうして読み出されたSMSデータと音楽パラメータとに基づき、様々なデータ操作処理を行ない、処理済みのSMSデータをSMSサウンド合成器110に与える。

【0050】SMSサウンド合成器110は、入力されたSMSデータに基づき、前述の米国特許第5,029,509号に示されたようなSMS合成技術に従ってサウンドの合成を行なうものである。このSMSサウンド合成器110の具体的構成としては、例えばその米国特許の第2図、第4図又は第5図に示されたような構成を参照することができる。しかし、便宜のために、SMSサウンド合成器110の基本構成例をブロック110内に概略的に示す。すなわち、入力されたSMSデータのうち、確定的成分に対応する線スペクトルデータ（周波数、マグニチュード、位相）が確定的波形発生部110aに入力され、これらに基づくフーリエ合成技術によって確定的成分に対応する波形が発生される。また、入力されたSMSデータのうち、ストカスティック成分に対応する残差スペクトルエンベロープがストカスティック波形発生部110bに入力され、このスペクトルエンベロープに対応するスペクトル特性を持つストカスティック波形が発生される。ストカスティック波形発生部110bは、例えば、ノイズ信号を残差スペクトルエンベロープに応じた特性でフィルタすることによりストカスティック波形を発生する。発生された確定的成分に対応する波形とストカスティック波形が加算器110cで加算され、望まれていたサウンドの波形信号が得られる。

【0051】再生処理部50では、合成すべきサウンドのピッチを、ユーザーの所望により自由に設定することが可能である。すなわち、ユーザーが所望のピッチを指定すると、これに応じて、SMSデータにおける周波数データを変更する処理を行ない、所望ピッチでのサウンド合成を可能にする。勿論、再生処理部50は、ユーザーによるリアルタイムでの発音指示に応じた1つのサウ

ンドの合成に限らず、例えば編集機器群13でプログラムされたデータに従って、複数のサウンドを、同時にまたは所定シーケンスで順番に、合成するよう処理することも可能である。ユーザーがリアルタイムで所望のボーカルフレーズに対応する制御パラメータを順次入力する、又は、プログラムされたデータに基づいて所望のボーカルフレーズに対応する制御パラメータを入力する、ことによって所望のボーカルフレーズの合成が可能である。

#### 【0052】再生処理部における処理例

再生処理部50において行なわれる各種処理の一例について図5を参照して説明する。図5は、再生処理部50において行なわれるすべての処理を紹介するものではなく、いくつかの代表的な処理について示している。図5に示された処理における特徴的事項は、データの補間と、音楽パラメータを考慮したSMSデータの再生である。データ補間を行なわない場合は、補間に関連する処理ステップを省略してよいのは勿論である。まず、データ補間を行なわない場合について説明する。その場合は、図5のステップ51～59が有効とされたと考えてよい。すなわち、現在発音すべきことが選択されている1つの音についてのみ処理が行なわれる。

#### 【0053】ステップ51：フレーム選択

ここでは、シンセサイザークロックに従って、現在のフレームが指定され、この現在フレームに対応するデータ（SMSデータと各種パラメータ）をデータメモリ100から取り出す。このフレーム選択処理のアルゴリズムは、シンセサイザークロックに従ってフレームを単純に進めることのみならず、前述のループ処理のために、ループエンドのフレームの次にループスタートのフレームに戻ることも行なうようにしてよい。

#### 【0054】ステップ52：データ変換

ここでは、データメモリ100から取り出された当該フレームの分析データ（SMSデータと音楽パラメータ）を、ユーザーによる制御に従って、変更する処理を行なう。例えば、所望のピッチがユーザーによって指示されると、それに応じて周波数データを変更する。あるいは、ユーザーによって所望のビブラートやトレモロが指示されると、それに応じて所定の音楽パラメータを変更する。こうして、ユーザーは、全ての分析データに関して、かつ全てのフレームにわたって、所望の制御を及ぼすことができる。このステップ52による変換を経由して各ステップ53～59に与えられるデータ名が例示的に図5に示されている。

【0055】ステップ53：ここでは、前記正規化されたピッチ関数を、全体平均ピッチによって演算し、正規化を解除したピッチ関数を得る。

ステップ54：ここでは、前記正規化されたマグニチュード関数を、全体平均マグニチュードによって演算し、正規化を解除したマグニチュード関数を得る。

#### ステップ55：周波数付加

ここでは、正規化されたSMSデータのうち、周波数データの値を、ピッチ関数を使用して正規化解除する。

#### 【0056】ステップ56：マグニチュード付加

ここでは、正規化されたSMSデータのうち、マグニチュードデータの値を、マグニチュード関数とチルトデータを使用して正規化解除する。SMSデータにおいて残差スペクトルエンベロープが正規化されている場合も、その正規化解除をここで行なう。

#### 10 ステップ57：ビブラート及びトレモロ付加

ここでは、ビブラートデータ及びトレモロデータを使用して、SMSデータにビブラート及びトレモロを付加する。

#### ステップ58：フォルマント付加

ここでは、フォルマントデータを使用して、SMSデータにフォルマントを付加する。

#### ステップ59：アーティキュレーション付加

ここでは、発生すべきサウンドにアーティキュレーションをつけるために、SMSデータに対して適宜のデータ処理を施す。

【0057】次に、データ補間について説明する。これは、発生すべきサウンドが、或る音（これを前音という）から別の音（これを現在音という）に移行するとき、スムーズな移行を可能にするための処理である。例えば、シンギング・ボイスを合成するとき有効である。このために、現在音の発生の始まりの適当な期間の間、前音の分析データ（SMSデータ及び各種パラメータ）もデータメモリ100から取り出すようにする。

#### 【0058】ステップ61：フレーム選択

30 ここでは、前音に関して適当なフレームのデータ（SMSデータと各種パラメータ）をデータメモリ100から取り出す。

#### ステップ62：データ変換

ここでは、ステップ52と同様に、当該フレームの分析データ（SMSデータと音楽パラメータ）を、ユーザーによる制御に従って、変更する。

#### ステップ65～71：補間

40 ここでは、SMSデータ及び各パラメータ毎に、前音のデータと現在音のデータとの間で、所定の補間特性に従って補間を行なう。この補間特性としては、例えば、クロスフェード補間のように前音のデータから現在音のデータへと時間的に滑らかに変化してゆくような特性を使用することができるが、その他の適宜の特性を使用してもよい。補間ステップ65～71における様々な補間演算パラメータを、ユーザーの制御に従って変更することができるようになっている。

【0059】〔各種のデータ処理機能の詳細〕次に、各種のデータ処理機能の詳細について説明する。以下では、各機能別に、分析から合成に至る処理が説明される。分析段階での処理は、SMSデータ処理部30（図

2, 図3)で実行され、合成段階での処理は再生処理部50(図4, 図5)で実行される。以下の説明では、各データ処理機能はSMSデータを対象にして施されるが、個別の各処理機能それ自体はSMSデータに限らず、その他のデータフォームからなる楽音データに適用可能であり、すべてのデータフォーム種類の楽音データに対する適用がクレームされた本願発明の範囲に含まれる。

#### 【0060】ーフォルマント抽出及び操作ー

この機能は、図3のステップ36及び図5のステップ58における処理に対応するものである。この機能に関わる発明の目的は、サウンドの線スペクトル(すなわちSMSデータにおける確定的表現物である周波数とマグニチュードつまり振幅のペアからなる1組の部分音成分)から、フォルマント構造(全体的なスペクトル特性)を抽出し、該サウンドの線スペクトルをフォルマント抽出物と残余のスペクトルに分離することにより、分析データの圧縮化を図ると共に、サウンド合成の際にフォルマントの変更等の制御を極めて容易にできるようにすることである。周知のように、ボーカルサウンドにおいては、そのボイスを特徴づけているフォルマントが存在するので、この機能はボーカルサウンドの分析及び合成において極めて有利である。

【0061】この機能に従うフォルマント抽出及び操作システムの全体的なブロック図を図6に示す。入力側に示されたSMS分析のステップと出力側に示されたSMS合成のステップは、前述のSMS分析器20とSMSサウンド合成器110による処理ステップに夫々対応している。前述のようにSMS分析によって得られるSMSデータは、周波数トラジェクトリ及びマグニチュードトラジェクトリと、ストカスティックエンベロープ(残差スペクトルエンベロープ)とを含む。このうち、ストカスティックエンベロープについてはこの機能に従う処理が施されず、確定的部分の分析結果つまり線スペクトルデータ即ち周波数トラジェクトリ及びマグニチュードトラジェクトリに対してこの機能に従う処理が施される。参考のために、フォルマントの特性を示している1フレーム分の確定的部分の分析結果つまり線スペクトルデータの一例を図7に示し、それに対応する1フレーム分のストカスティックエンベロープの一例を図8に示す。

【0062】図6において、ステップ80と81の処理は、図3のステップ36の処理に対応するものである。ステップ80では、1フレーム分の線スペクトルデータからフォルマントを抽出するための処理を行なう。すなわち、1セットの線スペクトルデータからフォルマントの山を検出することと、検出したフォルマントの山を適切な表現からなるパラメータで表現することとを行なう。このパラメータ表現は、前述したフォルマントデータに対応するものである。そして、各フレーム毎に、こ

のフォルマント抽出を行ない、フレーム毎のパラメータ表現つまりフォルマントデータを得る。こうして、フレーム毎に時変動可能である一連のフォルマントデータ(これをフォルマントトラジェクトリと呼ぶ)を得る。1セットの線スペクトルの中に複数のフォルマントがある場合、各フォルマント毎の連続的なフォルマントトラジェクトリがある。フォルマントデータのパラメータ表現の仕方として、指数近似を、ここではまず提案する。

【0063】通常、フォルマントは、パワースペクトルにおける三角形関数又はデシベルスペクトルにおける指数関数で記述することができる。デシベルスペクトルは人間の感覚に近いので、これを使用することは有益である。そこで、フォルマントの両側を夫々指数関数で近似することにする。そのために、フォルマントの各側毎に、そのスロープにフィットする最適の指数関数を見つけ出し、見つけ出した指数関数により該フォルマントを表現する。この最適の指数関数の見つけ出し方や、表現法には様々なバリエーションがあるであろう。その一例を、図9を参照して説明する。

【0064】この例では、次の4つの値によって1つのフォルマントを表現する。 $i$ は或る1つの時間フレームを特定するフレーム番号、 $i$ は或る1つのフォルマントを特定するフォルマント番号である。

- (1) 中心周波数  $F_i(i)$  :  $i$  番目のフォルマントの中心周波数を示すパラメータ
- (2) ピークレベル  $A_i(i)$  :  $i$  番目のフォルマントの中心周波数位置における振幅値を示すパラメータ
- (3) バンド幅  $B_i(i)$  :  $i$  番目のフォルマントのバンド幅を示すパラメータ
- (4) インターセクション  $E_i(i)$  :  $i$  番目のフォルマントとその隣の  $i+1$  番目のフォルマントとの交点を示すパラメータ

【0065】上記のうち最初の3つのパラメータは従来より知られたフォルマント表現であるが、最後のインターセクションパラメータは従来知られていなかったものである。これは、例えば、 $i$  番目のフォルマントとその隣の  $i+1$  番目のフォルマントとの交点に位置する1つの部分音成分すなわちスペクトラムの周波数を示すものである。ただし、最初の3つのパラメータに関しても、その求め方は、後述するように指数近似によって求める新規なものである。

【0066】ステップ80における処理手順を更に詳しく説明すると次の通りである。

- (1) フレーム  $i$  の各線スペクトルつまり部分音成分に対応するマグニチュードデータ  $a_n(i)$  の中からいくつかのローカル最大値を見つけ出す。ここで、前記式1のように、 $n$ は、 $n=0, 1, 2, \dots, N-1$  の夫々の値をとる変数であり、 $N$ は、そのフレームにおいて分析された線スペクトルつまり部分音成分の数である。
- (2) 見つけ出した個々のローカル最大値毎に、そのロー

カル最大値を取り囲んでいる2つのローカル最小値を、夫々見つけ出す。こうして見つけ出された1つのローカル最大値と、その両側の2つのローカル最小値は、1つの山を提示するものである。

【0067】(3) 各ローカル最大値とその両側の2つのローカル最小値とによって提示される各山から、前記パラメータ  $F_i$ ,  $A_i$ ,  $B_i$ ,  $E_i$  を夫々算出する。こうして、フレーム  $i$  についての各フォルマント  $i$  に対応するフォルマントデータ  $F_i$ ,  $A_i$ ,  $B_i$ ,  $E_i$  が得られる。

(4) 上記で求めたフレーム  $i$  についての各フォルマント  $i$  に対応するフォルマントデータを、個別のフォルマントトラジェクトリに割当てて、どのフォルマントトラジェクトリに割当ててかは、中心周波数が最も近いものを探し出して決定する。これによりフォルマントの連続性が確保される。過去のフォルマントトラジェクトリにおいて、所定の誤差範囲内で中心周波数が近いものがない場合は、そのフォルマントのために新たなフォルマントトラジェクトリを割当ててもよい。

【0068】上記(3)のステップにおける各パラメータ  $F_i$ ,  $A_i$ ,  $B_i$ ,  $E_i$  の算出アルゴリズムにつき、次に、説明する。上記(2)のステップにおいて1つのローカル最大値とその両側の2つのローカル最小値により1つの山が特定されると、それから、これに合う2つの側の指数関数を見つけ出さねばならない。この問題は、下記数2に示すような式によって数学的に公式化することができる。

【0069】

【数2】

$$e = \sum_{n=L1}^{n=Lr} \left\{ x^{-|F-f_n|} - \frac{a_n}{A} \right\}^2$$

ただし、 $n = L1, \dots, Lr$

【0070】ここで、 $F$  と  $A$  は未知数であり、求めるべきこのフォルマントにおける中心周波数とピークレベル振幅値である。 $L1$  と  $Lr$  は、2つのローカル最小値に対応する部分音成分の次数である。 $f_n$  と  $a_n$  は、この山の内側にある部分音成分  $i$  の周波数と振幅（つまりマグニチュード）である。 $x$  は、近似に使用する指数関数の底である。 $-|F-f_n|$  が、この指数関数の指数部である。 $e$  は、この指数関数と部分音成分との間の適合の誤差である。すなわち、上記式2は最小自乗近似法による誤差関数である。これにより、誤差  $e$  が最小となるような  $F$ ,  $A$ ,  $x$  を見つけ出す。これは最小限に見積もっても解くことが大変困難な問題である。しかし、本件での適用にあたっては、それほど厳密な適合が要求されないため、別の簡単な解決策を講じてよい。そこで、 $F$ ,  $A$ ,  $x$  を見つけ出すための、次のような、簡単なアルゴリズムを提案する。

【0071】その簡単なアルゴリズムとは、フォルマント周波数 ( $F$ ) とフォルマント振幅 ( $A$ ) を、ローカル最大値を精製することにより、得るものである。これは、その山における3つの最も高い振幅値について放物線の補間を行なうことによって行なう。その補間の結果得られる最大値の位置がフォルマント周波数 ( $F$ ) に相当し、その高さがフォルマント振幅 ( $A$ ) に相当する。フォルマントバンド幅  $B$  は、慣行的には、フォルマントの先端から  $-3$  dB 下がったところの帯域幅がそれに相当する。そのような値は指数関数の底  $x$  を記述する。それらは、下記式のような関係にある。

【0072】

【数3】

$$B = -2F \frac{\ln[(A-3)/A]}{\ln(x)}$$

【0073】すべての部分音成分について最も良く適合するバンド幅を持つフォルマント（指数関数）は次のようにして見つけ出される。まず、個々の部分音成分  $n$  について下記式による指数関数の値  $x_n$  を夫々求める。

【0074】

【数4】

$$x_n = e^{\frac{\ln(a_n/A)}{F - f_n}}$$

【0075】それから、各  $n$  に対応する上記指数関数値  $x_n$  を上記数3の式の  $x$  に代入して、夫々に対応する仮のバンド幅  $B_n$  を夫々求める。こうして求めたそのフォルマントの各仮のバンド幅  $B_n$  を下記式のように平均化する。

【0076】

【数5】

$$B = \frac{1}{Lr-L1} \sum_{n=L1}^{n=Lr} B_n \quad \text{ただし、} n = L1, \dots, Lr$$

【0077】この平均バンド幅  $B$  が、そのフォルマントのバンド幅として使用され、フォルマントとして使用された指数関数を記述するものとなる。 $i$  番目のフォルマントとその隣の  $i+1$  番目のフォルマントとの交点を示すインターセクションパラメータ  $E_i$  は、そのフォルマント  $i$  における右側のローカル最小値の周波数を用いる。

【0078】図6に戻ると、ステップ81では、上記のように抽出した1フレームのフォルマントデータを使用して、そのフレームについての1組の部分音成分からフォルマント構造を引算する。フォルマント構造は、フォルマントの形状を示す相対値であると考えてよい。1組の部分音成分つまり線スペクトルからフォルマント構造を引算することは、フォルマントによる変化分を差し引いて、1組の部分音成分つまり確定的成分の線スペクト

ルを平坦化することである。従って、このステップ81の処理の結果得られる、確定的成分の線スペクトルデータは、例えば図10のように、平坦化されたスペクトル構造を持つものとなる。

【0079】この手法の一例を示すと、1フレームのすべてのフォルマントデータに基づき、該フレームのすべてのフォルマントを記述する関数を発生し、この関数が0平均を持つようにその振幅値を正規化する。このように正規化されたフォルマント関数は、フォルマント構造を示している。そして、そのフレームについての1組の部分音成分における個々の部分音成分毎に、そのマグニチュード値からその周波数位置に対応する正規化フォルマント関数の振幅値を引算する。勿論、その他の手法も可能である。

【0080】ステップ82の処理は、図5のステップ52, 62, 71の処理に対応するものである。すなわち、上記のように抽出されたフォルマントデータをユーザーの制御によって自由に変更する処理が行なわれる。ステップ83の処理は、図5のステップ58の処理に対応するものである。すなわち、上記のように変更が適宜加えられたフォルマントデータを確定的成分の線スペクトルデータに付加し、確定的成分の線スペクトルデータにフォルマント特性を持たせる。

【0081】このフォルマント操作によれば、ユーザーは4つのパラメータF, A, B, Eを望みに応じて制御することにより、フォルマントを自由に制御することができる。これらの4つのパラメータF, A, B, Eは、フォルマントの特性/形状に直接対応しているので、フォルマント操作/制御が非常にし易いものとなる、という利点がある。また、フォルマントの分析/抽出についても、上記で提案した方法は、従来知られたLPCのような自乗近似法に比べて簡単であり、計算も能率的に行なうことができる、という利点がある。

【0082】一フォルマント抽出及び操作の別の例ー図11は、フォルマント抽出及び操作システムの別の例を示す全体的なブロック図である。ここでは、フォルマントを抽出するためのステップ80aが図6のステップ80と相違しており、他は同じであってよい。このシステムでは、フォルマントは、デシベルスペクトルにおける二等辺三角形関数で近似される。デシベルスペクトルは人間の感覚に近いので、これを使用することは有益である。フォルマントのスロープにフィットする最適の二等辺三角形関数を見つけ出し、見つけ出した二等辺三角形関数により該フォルマントを表現する。この最適の二等辺三角形関数の見つけ出し方や、表現法には様々なバリエーションがあるであろう。その一例を、図12を参照して説明する。

【0083】この例では、次の3つの値によって1つのフォルマントを表現する。 $i$ は或る1つの時間フレームを特定するフレーム番号、 $l$ は或る1つのフォルマント

を特定するフォルマント番号である。

(1) 中心周波数 $F_i(l)$ ： $i$ 番目のフォルマントの中心周波数を示すパラメータ

(2) ピークレベル $A_i(l)$ ： $i$ 番目のフォルマントの中心周波数位置における振幅値を示すパラメータ

(3) スロープ $S_i(l)$ ： $i$ 番目のフォルマントのスロープ(二等辺三角形の辺の傾き)を示すパラメータ

上記のうち最初の2つのパラメータは従来より知られたフォルマント表現であるが、最後のスロープパラメータは従来知られていなかったものであり、これは、従来より知られたバンド幅に置き換わる新規なものである。このスロープをバンド幅に変換することは容易に行なえる。

【0084】ステップ80aにおける処理手順を更に詳しく説明すると次の通りである。

(1) 山の検出：フレーム $l$ の各線スペクトルつまり部分音成分に対応するマグニチュードデータ $a_n(l)$ の中からいくつかのローカル最大値つまりピークを見つけ出す。また、見つけ出した個々のローカル最大値毎に、そのローカル最大値を取り囲んでいる2つのローカル最小値つまり谷を、夫々見つけ出す。こうして見つけ出された1つのローカル最大値と、その両側の2つのローカル最小値は、1つの山を提示するものである。このような山検出の一例を図13に示す。

(2) 三角形適合：各ローカル最大値とその両側の2つのローカル最小値とによって提示される各山から、三角形近似によって、前記パラメータ $F_i$ ,  $A_i$ ,  $S_i$ を夫々算出する。こうして、フレーム $l$ についての各フォルマント $i$ に対応するフォルマントデータ $F_i$ ,  $A_i$ ,  $S_i$ が得られる。

【0085】(3) 上記で求めたフレーム $l$ についての各フォルマント $i$ に対応するフォルマントデータを、個別のフォルマントトラジェクトリに割当てて。どのフォルマントトラジェクトリに割当ててかは、中心周波数が最も近いものを探し出して決定する。これによりフォルマントの連続性が確保される。前述と同様に、過去のフォルマントトラジェクトリにおいて、所定の誤差範囲内で中心周波数が近いものがない場合は、そのフォルマントのために新たなフォルマントトラジェクトリを割当ててもよい。図16は、フォルマントトラジェクトリの様子を模式的に示すマップである。

【0086】上記(1)のステップにおける山検出について更に説明する。一例として、隣接する3つの部分音成分のマグニチュードつまり振幅値 $a_{-1}$ ,  $a_0$ ,  $a_1$ が下記式を満足するとき、その中央のマグニチュード $a_0$ に対応する部分音成分をローカル最大値として検出するようにしてよい。

【0087】

【数6】

$$a_{-1} \leq a_0 \leq a_1$$

【0088】そして、ローカル最大値の両隣の谷を同じ様な手法でローカル最小値として検出する。次に、上記(2)のステップにおける各パラメータ  $F_i$ ,  $A_i$ ,  $S_i$  の算出アルゴリズムにつき、説明する。まず、中心周波数  $F_i$  は、前述と同様に、その山における3つの最も高い振幅値について放物線の補間を行なうことによって見つけ出す。このためのアルゴリズムとしては、下記式を用いることができる。

【0089】

【数7】

$$d = \frac{0.5(a_{-1} - a_1)}{a_{-1} - 2a_0 + a_1}$$

【数8】

ここで、

$$d < 0 \text{ ならば、} F_i = f_0 + d(f_0 - f_{-1})$$

$$d \geq 0 \text{ ならば、} F_i = f_0 + d(f_1 - f_0)$$

【0090】ここで、 $f_{-1}$ ,  $f_0$ ,  $f_1$  は、前述の各マグニチュード  $a_{-1}$ ,  $a_0$ ,  $a_1$  に対応する隣接する3つの部分音成分の周波数である。 $d$  は、そのうち中央の周波数  $f_0$  からの中心周波数  $F_i$  の距離である。まず数7の式により  $d$  を求め、求めた  $d$  を数8の式に適用して  $F_i$  を求める。

【0091】次に、各部分音成分  $n$  を中心周波数  $F_i$  からの隔たりに応じた相対値 ( $x_n$ ,  $y_n$ ) に置き換えたデータセットを作成する。 $x_n$  は周波数の相対値であり、下記式で得られる。

$$\text{【数9】 } x_n = |F_i - f_n|$$

$f_n$  は各部分音成分  $n$  の周波数である。数9の式では差の絶対値が周波数の相対値  $x_n$  となっているため、図14に模式的に示すように、すべての  $x_n$  が  $F_i$  の片側に \*

$$D_x = \sum_{n=L1}^{n=Lr} x_n$$

$$D_{xx} = \sum_{n=L1}^{n=Lr} x_n^2$$

【0099】こうして得られた上記関数の傾き  $S_i$  は、三角形の右側のスロープに対応するものである。その左側のスロープは、 $-S_i$  である。また、関数のオフセット値  $A_i$  は、フォルマントのピークレベルに対応する。以上により、フォルマントに最も適合する二等辺三角形近似を定義する3つのパラメータ  $F_i$ ,  $A_i$ ,  $S_i$  を得ることができる。図15はそのようなフォルマントの二等辺三角形近似を示すものである。

【0100】前述のように、フォルマントのバンド幅B

\* くるように折り返されることになる。 $y_n$  は、各相対周波数  $x_n$  に対応する部分音成分  $n$  の振幅であり、これは下記のように各部分音成分  $n$  のマグニチュード  $a_n$  にそのまま対応している。

【0092】

$$\text{【数10】 } y_n = a_n$$

【0093】こうして、三角形適合プログラムを、単純な線適合プログラムに変換することができる。すなわち、下記のような1次関数  $y$  を用いて  $A_i$  と  $S_i$  を見つけ出すことができる。

$$\text{【数11】 } y = A_i + S_i \cdot x$$

この数11の式の  $x$  と  $y$  に、上記データセット ( $x_n$ ,  $y_n$ ) を夫々代入し、下記の最小自乗近似式に従い、誤差  $e$  を最小にするような  $A_i$  と  $S_i$  を見つけ出す。

【0094】

【数12】

$$e = \sum_{n=L1}^{n=Lr} (y_n - A_i - S_i x_n)^2$$

ただし、 $n = L1, \dots, Lr$

【0095】 $L1$  と  $Lr$  は、2つのローカル最小値つまり谷に対応する部分音成分の次数である。この解  $A_i$ ,  $S_i$  は下記式のように得られる。

【0096】

【数13】

$$A_i = \frac{D_{xx} D_y - D_x D_{xy}}{D_x^2 - D_{xx}}$$

$$S_i = \frac{D_{xy} - D_x D_y}{D_x^2 - D_{xx}}$$

【0097】ここで、各導関数  $D_x$ ,  $D_y$ ,  $D_{xx}$ ,  $D_{xy}$  は次の通りである。

【0098】

【数14】

$$D_y = \sum_{n=L1}^{n=Lr} y_n$$

$$D_{xy} = \sum_{n=L1}^{n=Lr} x_n y_n$$

$i$  は、慣行的には、フォルマントの先端から  $-3$  dB 下がったところの帯域幅がそれに相当するので、フォルマント中心周波数  $F_i$  とスロープ  $S_i$  とに基づき、下記式により容易に求めることができる。

【0101】

【数15】

$$B_i = 2 \left( \frac{-3}{S_i} + F_i \right)$$

【0102】スロープパラメータ  $S_i$  はそのままフォル



マント変更ステップ83に与えてもよいし、バンド幅パラメータに変換してからフォルマント変更ステップ83に与えるようにしてもよい。なお、変形例として、二等辺三角形近似に限らず、その他の不等辺三角形近似により各側のスロープを別々に近似することにより、フォルマントの三角形近似を行なうようにすることができる。

【0103】このフォルマント操作によれば、ユーザーは3つのパラメータF、A、Sを望みに応じて制御することにより、フォルマントを自由に制御することができる。これらの3つのパラメータF、A、Sは、フォルマントの特性/形状に直接対応しているため、フォルマント操作/制御が非常にし易いものとなる、という利点がある。また、フォルマントの分析/抽出についても、上記で提案した方法は、従来知られたLPCのような自乗近似法に比べて簡単であり、計算も能率的に行なうことができる、という利点がある。また、三角形近似によりフォルマントデータを抽出するので、抽出のための計算のアルゴリズムが非常に簡単であるという利点がある。更に、二等辺三角形近似によりフォルマントの分析/抽出を行なうことにより、片側のスロープのみを計算すればよいことになるので、アルゴリズムを更に簡単化することができるという利点を持つ。

#### 【0104】-ビブラート分析及び操作-

ビブラートは、各部分音成分毎に、その周波数トラジェクトリの時間関数を分析することによって検出する。図17は、ビブラート分析システムの一例を示す全体的なブロック図である。これは、図3のステップ37の処理に対応している。ビブラート分析は各部分音成分毎に行なうので、この分析システムの入力は、或る1つの部分音成分の周波数トラジェクトリであり、これは、各時間フレーム毎の周波数を示す時間関数である。容易に理解できるように、この周波数の時間関数が、ビブラートとみなすことができる周期で時変動していれば、その時変動成分をビブラートとして検出することができる。従って、周波数トラジェクトリの時間関数における低周波数の時変動成分を検出することによってビブラートの検出を行なうことができる。そのために、図17では、高速フーリエ変換技術を使用してビブラート分析を行なうようにしている。

【0105】まず、ゲート90では、分析対象である1つの周波数トラジェクトリの時間関数を入力し、ビブラート分析用の所定の時間窓信号によってゲートする。この時間窓信号は、隣接するフレームにおいてそのフレームサイズが所定割合で（例えば3/4づつ）オーバーラップするように、周波数トラジェクトリの時間関数をゲートする。なお、ここでいうフレームとは、前述のSMSデータにおける時間フレームとは異なるものであり、それよりもかなり長い時間に対応している。例えば、時間窓信号によって設定する1つのフレームが0.4秒の時間長を持つとすると、オーバーラップ割合が3/4である

とすると、隣接するフレーム間では、0.1秒の時間差を持つ。つまり0.1秒ごとの時間レートでビブラート分析がなされることになる。

【0106】ゲートされた信号は、直流除去器91に入力され、直流分を除去する。これは、例えば、そのフレーム内の関数値の平均値を求め、この平均値を直流分として除去する、すなわち各関数値から平均値を引算することによって行なうことができる。それから、高速フーリエ変換器(FFT)92に入力され、そのスペクトル分析がなされる。こうして周波数トラジェクトリの時間関数が時間窓信号によって複数のフレームに分割され、各フレーム毎にその交流的成分についてのFFT分析が行なわれる。FFT92による分析出力は複素スペクトルであるから、次の直交-極座標変換器93でマグニチュードスペクトル及び位相スペクトルに変換する。こうして得られたマグニチュードスペクトルがピーク検出及び補間部94に与えられる。

【0107】上記マグニチュードスペクトルの一例をエンベロープによって示すと図18のようである。オリジナルサウンドにビブラートがある場合は、ビブラートの可能性のある所定の周波数領域、例えば4Hz乃至12Hzの領域に、図示のようなピークが生じる。そこで、この領域におけるピークを検出し、その周波数位置をビブラートレートとして検出する。そのための処理をピーク検出及び補間のためのステップ94で行なう。このピーク検出及び補間のためのステップ94における処理例は次の通りである。

【0108】(1) まず、与えられたマグニチュードスペクトルのうち、ビブラートの可能性のある所定の周波数領域において振幅の最大値、つまりローカル最大値を検出する。図20は、ビブラートの可能性のある所定の周波数領域を拡大して示しており、kがローカル最大値のスペクトルに相当し、k-1とk+1がその両隣のスペクトルに相当する。

(2) 次に、上記ローカル最大値とその両隣のスペクトルの振幅値を通る放物線を補間する。図20におけるカーブPIは、この補間によって得た放物線を示す。

(3) 次に、補間によって得た放物線カーブPIにおける最大値を特定し、この最大値に対応する周波数位置をビブラートレートとして検出すると共に、この補間された最大値をビブラート幅として検出する。音楽パラメータとして抽出されるビブラートデータは、これらのビブラートレートとビブラート幅とからなっている。このビブラートデータの抽出が各フレーム毎に行なわれるので、時変動するビブラートデータの抽出が可能であることが理解できるであろう。

【0109】図17に戻ると、ステップ95では、直交-極座標変換器93で得たマグニチュードスペクトルから、ステップ94で検出したビブラート成分を引算する処理を行なう。ここでは、検出したビブラートの山の両

側の境界つまり2つの谷を見つけ出し、図19に示すように、この間を直線補間してビブラート成分の山を取り除く。図19は、このステップ95で処理されたマグニチュードスペクトルの一例を模式的に示している。

【0110】次に、ビブラート成分が除去されたマグニチュードスペクトルデータと、直交一極座標変換器93で得た位相スペクトルデータとを、極一直交座標変換器96に入力し、これらを複素スペクトルデータに変換する。それから、この複素スペクトルデータを逆FFT97に入力し、時間関数を発生する。この出力を直流加算部98に与え、前記直流除去器91で除去した直流分を再加算し、ビブラート成分が除去された1フレーム分の周波数トラジェクトリの時間関数を生成する。こうして、ビブラート成分が除去された1フレーム分の周波数トラジェクトリを各フレーム毎に連結して、その部分音成分に対応する一連の周波数トラジェクトリを作成する。その際に、前述のようにオーバーラップしたフレームの時間だけ、データを重複して連結するものとする。データ重複部分の連結の仕方としては、平均値を採用するのがよいと思われるが、その他の適宜の補間であってもよい。また、オーバーラップ部分において或る1つのフレームのデータのみ選択し、他を切り捨ててもよい。このようなオーバーラップ部分についての処理は、前記検出したビブラートレート及びビブラート幅のデータについても適宜行なってもよい。

【0111】図21は、ビブラート合成アルゴリズムの一例を示す全体的なブロック図である。ステップ85、86の処理は、図5のステップ52、62、69の処理に対応するものである。すなわち、上記のように抽出されたビブラートレート及びビブラート幅のデータを、ユーザーの制御によって自由に変更する処理が行なわれる。ステップ87及び88の処理は、図5のステップ57の処理に対応するものである。ステップ87では、上記のように変更が適宜加えられたビブラートレート及びビブラート幅のデータに基づき、ビブラート信号を例えば正弦波関数で発生する。ステップ88では、このビブラートレートとビブラート幅に対応する正弦波関数によって、SMSデータにおける対応する周波数トラジェクトリにおける周波数値を変調する演算を行なう。これにより、ビブラート付与された周波数トラジェクトリが得られる。

【0112】以上の説明では、各部分音成分毎に別々に、ビブラートデータを抽出し、制御若しくは変更し、かつ、ビブラート合成を行なうようにしている。しかし、各部分音成分毎にビブラートレートを異ならせる必要はないので、基本波成分から抽出したビブラートレート、あるいは低次のいくつかの部分音成分から抽出したビブラートレートの平均値、を各部分音成分に共通に使用するようにしてもよい。ビブラート幅についても同様に所定のものを各部分音成分に共通に使用するようにし

てよい。

#### 【0113】ートレモロの抽出及び操作ー

トレモロは、各部分音成分毎に、そのマグニチュードトラジェクトリの時間関数を分析することによって検出する。トレモロは振幅のビブラートであるといえるので、前述したビブラートの分析及び合成のアルゴリズムと同じものをそっくり利用することができる。ビブラートとの違いは、トレモロにおいては分析及び合成の対象がSMSデータにおけるマグニチュードトラジェクトリである、という点だけである。すなわち、図17乃至図21を参照して説明したのと同様の分析及び合成のアルゴリズムをマグニチュードトラジェクトリに対して適用することにより、トレモロの分析及び合成を行なうことができる。従って、図17乃至図21における“周波数トラジェクトリ”を“マグニチュードトラジェクトリ”と読み変えることにより、トレモロの分析及び合成のための実施例を提示することができる。トレモロデータとしては、トレモロレートとトレモロ幅とからなるパラメータが得られることになる。

【0114】同様に、SMSデータにおけるストカスティック成分に関しても、トレモロと同様の振幅の周期的変動を分析し、これを制御若しくは変更し、かつ、合成するようにすることができる。SMSデータにおけるストカスティック成分に対応する残差スペクトルエンベロープデータの1つとして、該スペクトルエンベロープの全体的ゲインを示すデータがあり、これをストカスティックゲインと呼ぶ。各時間フレーム毎の一連のストカスティックゲインをストカスティックゲイントラジェクトリと呼ぶ。ストカスティックゲイントラジェクトリはストカスティックゲインの時間関数である。従って、このストカスティックゲインの時間関数を前記ビブラート又はトレモロの場合と同様のアルゴリズムによって分析し、その分析結果を利用した制御と合成が可能である。また、分析を省略し、確定的成分のマグニチュードトラジェクトリの分析によって得たトレモロデータを使用してストカスティックゲインの制御と合成を行なってもよい。上述のようなビブラートあるいはトレモロの分析と制御及び合成の手法は、SMS合成技術に限らず、他の加算的楽音合成技術にも応用可能である。

#### 【0115】一音楽サウンドにおけるスペクトルチルト制御ー

図22は、この実施例に従うスペクトルチルト制御のための分析及び合成のアルゴリズムを示す。ステップ120～123は分析アルゴリズムに対応しており、SMSデータ処理部30(図2)で実行される。ステップ124、125は合成アルゴリズムに対応しており、再生処理部50(図4)で実行される。

【0116】スペクトルチルトの分析：まず、スペクトルチルトの分析について説明する。スペクトルチルト分析は、確定的成分に関して行なう。図23は、確定的成

10

20

30

40

50

35

分の線スペクトル例と、そこから分析した1直線状のスロープからなるスペクトルチルトラインの一例を示している。分析したスペクトルチルトラインは太い実線で示している。このスペクトルチルトラインの原点は、確定的成分の線スペクトルにおける最も低い周波数を持つ第1の部分音成分のマグニチュードレベル値である。そして、残りの全ての部分音成分のマグニチュード値を概ね近似することのできるような最適の傾きラインを見つけ出す(ステップ120)。これはラインフィッティングの問題であるから、スペクトルチルトのスロープbは次式によって計算できる。

【0117】

【数16】

$$b = \frac{\sum_{i=1}^{i=N-1} (x_i / x_0) \cdot (y_i - y_0)}{\sum_{i=1}^{i=N-1} (x_i / x_0)^2}$$

【0118】ここで、iは部分音番号、Nは部分音の合計数、xは各部分音の周波数、yは各部分音のマグニチュード値である。特定のSMS時間フレームについての平均マグニチュードmagは次式により計算できる。

【0119】

【数17】

$$mag = \frac{1}{N} \sum_{i=0}^{i=N-1} y_i$$

【0120】これらの計算により、スペクトルチルト(b)と平均マグニチュードmagのデータペアを各SMS時間フレーム毎に得ることができる。次に、各フレーム毎の平均マグニチュードmagの平均値すなわち全体平均マグニチュードAvgMagを計算する。そして、次式によってこれらの値の相関を求める(ステップ121)。

【0121】

【数18】

$$corr = \frac{\sum_{i=1}^{i=M} b_i \cdot (mag_i - AvgMag)}{\sum_{i=1}^{i=M} b_i^2}$$

【0122】ここで、iはSMS時間フレーム番号、MはSMS時間フレームの合計数である。この相関データcorrは、各フレームi毎の平均マグニチュードmag<sub>i</sub>に対する全体平均マグニチュードAvgMagの差(mag<sub>i</sub> - AvgMag)と、各フレームi毎のスペクトルチルトb<sub>i</sub>との相関を示すものである。すなわち、相関データcorrは、各フレーム毎のスペクトルチルトデータbを、そのフレームの平均マグニチュード

36

magに対する全体平均マグニチュードAvgMagの差(mag - AvgMag)に相関するデータとして正規化したものである。式18から容易に理解できるように、仮に、全フレームiのスペクトルチルトb<sub>i</sub>が等しいとすると、個別サンプルmag<sub>i</sub>とそれらの平均値AvgMagとの差(mag<sub>i</sub> - AvgMag)の合計は0に収束するので、相関データcorrは0である。ここから理解できることは、相関データcorrは、各フレームのスペクトルチルトb<sub>i</sub>の相互関係を、そのフレームi毎の平均マグニチュードmag<sub>i</sub>に対する全体平均マグニチュードAvgMagの差(mag<sub>i</sub> - AvgMag)をパラメータとして、示している基準値若しくは正規化値である、ということである。

【0123】以上によって求められた相関データcorrが、スペクトルチルトに関する唯一の音楽パラメータ、つまりチルトファクターである。ユーザーは、このチルトファクターつまり相関データcorrを変更制御することにより、合成するサウンドの明るさ等の表情を自由に制御することができる。なお、チルト分析にあたっては、確定的成分における全ての部分音を考慮にいれる必要はなく、適宜省略してよい。例えば、上記式16の分析式に算入する部分音成分を定義するために、或るスレシヨルド値を設定し、このスレシヨルド値以上のマグニチュードを持つ部分音成分を算入して分析を行なうようにしてよい。また、所定の高い周波数(例えば8000Hz)以上の高い周波数の部分音成分も上記式16の分析式に算入しないようにし、チルト分析にあたっての不安定要素を排除してよい。勿論、上記分析の結果得たスロープと実際の各部分音のマグニチュードとを照合し、あまりにもかけ離れているものがある場合は、それを除外して、もう一度分析をやり直すようにしてもよい。

【0124】スペクトルチルトによる正規化：次に、上記のように求めたスペクトルチルト分析データを使用して、SMSデータの確定的成分のマグニチュード値を正規化する処理を行なう。ここでは、各フレーム毎の確定的成分の線スペクトルが、見掛け上共通のスペクトルチルトを持つかのように、かつ、全体平均マグニチュードAvgMagに関して、夫々の部分音のマグニチュード値を正規化する。そのために、下記式に従って、各部分音成分毎に差分値diffを計算する(ステップ122)。

【0125】

【数19】

$$diff = corr \cdot (AvgMag - mag) \cdot (x_i / x_0)$$

【0126】ここで、magはそのフレームの平均マグニチュード、x<sub>0</sub>はそのフレームにおける第1の部分音の周波数、x<sub>i</sub>はこの計算の対象となっている部分音iの周波数である。それから、各部分音毎に計算した上記

差分値  $d i f f$  を、対応する部分音のマグニチュード値に加算し、正規化したマグニチュード値を求める（ステップ123）。

【0127】スペクトルチルト合成：前述のように、ユーザーは、分析されたチルトファクターつまり相関データ  $c o r r$  を自由に変更制御することができる（ステップ

$$d i f f = c o r r' * (n e w m a g - A v g M a g) * (x i / x 0)$$

【0128】ここで、 $c o r r'$  はユーザーによる変更制御処理を経たチルトファクターつまり相関データ、 $n e w m a g$  はそのフレームの平均マグニチュードであり、合成の際に適宜の処理が施されているかもしれないもの、 $x 0$  はそのフレームにおける第1の部分音の周波数、 $x i$  はこの計算の対象となっている部分音  $i$  の周波数である。これにより、チルトファクター  $c o r r'$  を考慮に入れた合成用の差分値  $d i f f$  が各部分音毎に求められる。この合成用の差分値  $d i f f$  を対応する部分音のマグニチュード値に加算することにより、望みの修正されたスペクトルチルトによって制御された線スペクトルデータを得る（ステップ125）。この修正された線スペクトルデータを含むSMSデータに基づき、後に、SMSサウンド合成器110（図4）でサウンド合成がなされる。従って、ユーザーによるチルトファクターつまり相関データ  $c o r r$  の変更制御に応じて、明るさ等の表情が自由に制御されたサウンドが合成される。

【0129】容易に理解できるように、もし、スペクトルチルトが時変動しない簡略化された制御を行なう場合は、相関  $c o r r$  の算出等の面倒な演算は省略できるであろう。つまり、分析されたスペクトルチルトデータ  $b$  をそのままユーザーによって自由に制御し、制御されたスペクトルチルトデータに基づきサウンド合成の際に線スペクトルのチルトを制御するようにしてもよい。発明の本質は、スペクトルのチルトを抽出し、これを制御することにより合成すべきサウンドの制御を行なう点にあるのであるから、そのような簡略化されたチルト分析と合成も、本発明の範囲に含まれると理解すべきである。このスペクトルチルト制御もまた、他の制御と同様に、SMS技術に限らず、他の部分音加算合成技術においても適用可能である。

#### 【0130】—サウンドの時間変更—

この技術の目的は、SMS技術によって表現されたサウンドの発音時間長を長くしたり又は短くしたりする制御を行なうことである。発音時間長を長くすることは、サンプラーにおけるルーピング技術で知られているように、サウンドの或る部分を切り出し、これを繰返しつつ合わせることによって行なう。発音時間を短くすることは、サウンドから適切に選んだセグメントを取り除くことによって行なう。以下で述べる例では、ループポイントを設定するために、ビブラートサイクルの境界を見つけ出すようにしたことを特徴としている。

\* プ124）。サウンド合成に際しては、各部分音成分のマグニチュード値をチルトファクターによって制御する処理を行なう。そのために、下記式に従って、各部分音毎に合成用の差分値  $d i f f$  を計算する。

【数20】

【0131】図24は、この実施例に従う時間変更のための分析及び合成のアルゴリズムを示す。ステップ130、131、132は分析アルゴリズムに対応しており、SMSデータ処理部30（図2）で実行される。ステップ133、134、135は合成アルゴリズムに対応しており、再生処理部50（図4）で実行される。ステップ130、131、132による分析アルゴリズムによれば、オリジナルサウンドのビブラートサイクルの境界を見つけ出す処理を行なう。そのために、ビブラートの特徴が現われやすい低次の部分音成分のいくつかの周波数トラジェクトリを対象にして分析を行なう。この例では、第1の部分音成分すなわち基本波と第2の部分音成分すなわち第1ハーモニクの2つの周波数トラジェクトリに関して、夫々分析を行なう。

【0132】まず、ステップ130では、分析しようとする音の中央当たりにおいて、基本波の周波数トラジェクトリと第1ハーモニクの周波数トラジェクトリから、その周波数が最も高いローカル最大値を探し出す。これを最初のローカル最大値とする。具体的には、分析しようとする音の中央当たりの所定時間範囲内において、基本波の周波数トラジェクトリと第1ハーモニクの周波数トラジェクトリの夫々につき、7フレーム分の周波数の平均値を順次に作成し、そのファイルを作成する（7ポイント平均値ファイルの作成）。こうして、作成した各トラジェクトリの7ポイント平均値ファイルを比較参照して、基本波と第1ハーモニクの両方に関して生じている最も高いローカル最大値を探し出す。こうして、探し出したローカル最大値の位置と値を、最初のローカル最大値としてリストに入れる（最初のローカル最大値の検出）。仮にオリジナルサウンドにビブラートがなかったとしても、このようなローカル最大値の検出は可能である。なお、SMS時間フレームのレートを100Hzとすると、そのような7ポイントつまり7フレームの長さは0.07秒である。

【0133】次に、ステップ131では、上記のように見つけ出した最初のローカル最大値の位置を基に、さらにサーチを進め、その両側において周波数が最小である2つのローカル最小値を探し出して、上記最初のローカル最大値のリストに加える。それから、更に時間進行方向にサーチを進め、音の終了近くまでに、いくつかのローカル最大値とローカル最小値のペアを探し出し、上記リストに時間順に加える。こうして、探し出したすべて

のローカル最大値とローカル最小値つまり極値の値と位置が上記リスト（つまり極値リスト）に時間順に記憶される。

【0134】具体的には、まず、上記各トラジェクトリの7ポイント平均値ファイルにおいて最初のローカル最大値の位置から時間進行方向にサーチを進め、基本波と第1ハーモニックの両方に関して生じている周波数が最も低いローカル最小値（右のローカル最小値）を探し出す。このとき、必要に応じて、分析対象範囲を時間進行方向に広げて、前記7ポイント平均値ファイルにファイルする各トラジェクトリの7ポイント平均値データを追加作成する。こうして、探し出した右のローカル最小値の位置と値を、上記極値リストにおいて最初のローカル最大値の右隣に記憶する（右ローカル最小値検出）。

【0135】次に、上記各トラジェクトリの7ポイント平均値ファイルにおいて最初のローカル最大値の位置から時間逆行方向にサーチを進め、基本波と第1ハーモニックの両方に関して生じている周波数が最も低いローカル最小値（左のローカル最小値）を探し出す。このときも、必要に応じて、分析対象範囲を時間逆行方向に広げて、前記7ポイント平均値ファイルにファイルする各トラジェクトリの7ポイント平均値データを追加作成する。こうして、探し出した左のローカル最小値の位置と値を、上記極値リストにおいて最初のローカル最大値の左隣に記憶する（左ローカル最小値検出）。

【0136】次に、分析対象範囲を時間進行方向に、音の終了近くまで、広げて、前記7ポイント平均値ファイルにファイルする各トラジェクトリの7ポイント平均値データを追加作成する。それから、前述と同様に、各トラジェクトリの7ポイント平均値ファイルにおいて時間進行方向にサーチを進め、基本波と第1ハーモニックの両方に関して生じている周波数の極値（ローカル最大値又はローカル最小値）を順次検出し、これらの位置と値を上記極値リストに時間順に記憶する。こうして作成された極値リストにリストされた各極値のいくつかは、ビブラートサイクルのピークと谷であると推定することができる。なお、極値の位置データとは、時間に対応するデータである。次のステップ132では、上記ステップ131でリストした極値データを検討し、ビブラートサイクルのピークと谷であると推定される極値データを残し、他を削除するための編集処理を行なう。

【0137】具体的には、次のように処理する。まず、リストした極値データにおいて見られるビブラートサイクルが、所定のビブラートレートの範囲内に納まっているかを調べる。すなわち、極値リストにおける或る最大値と或る最小値の時間差が所定の時間範囲内に納まるかを、全ての最大値と最小値のペアにつき、夫々調べる。所定の時間範囲の一例を示すと、最大で0.15秒、最小で0.05秒である。こうして、所定の時間範囲に納まっていない最大値と最小値のいくつかのペアを見つけ

出すことができるであろう。これらの各ペアのうち少なくとも一方は、ビブラートの最大値又は最小値に対応していないものである。こうして、調べた結果、その時間差が所定の時間範囲に納まっている各極値ペアを、保存すべきものとして、マークする。ところで、上記所定時間範囲はむしろ広めに設定してあるので、有効なビブラート極値がマークされないことは有りえない。しかし、そのために、実際のビブラートを示している極値よりも多くの極値がマークされてしまう可能性がでてくる。なお、ここでマークされなかった極値は、以後の処理では全て無視される。

【0138】次に、リストに保存された各極値ペアにおいて、最小値から最大値に向かうアップスロープの時間間隔と、最大値から最小値に向かうダウンスロープの時間間隔を夫々算出する（図25参照）。そして、夫々のアップスロープ時間間隔の平均値と、夫々のダウンスロープ時間間隔の平均値を計算する。それから、各極値ペア毎のアップスロープ時間間隔と上記アップスロープ平均値との関係、及び各極値ペア毎のダウンスロープ時間間隔と上記ダウンスロープ平均値との関係、を夫々調べ、夫々の時間間隔が平均値に対して所定の誤差限界内に収まっているかを調べる。例えば、この誤差限界としては、平均値の20%としてよい。この誤差限界内に収まっている各極値ペアを、保存すべきものとして、マークする。最初と最後の極値を除く各極値は、アップスロープとダウンスロープに関して合計2回の検査を受けることになる。どちらかの検査が合格であれば、その極値を保存すべきことがマークされることになる。

【0139】以上の処理を経た結果として極値リストに保存された極値がビブラートの最大値及び最小値として推定できるものである。ルーピングのためにつなぎ波形として使用するセグメントは、2つの最大値又は2つの最小値の間の波形とする。そのために、少なくとも3つの極値がリストに保存されていなければならない。もし、2以下の極値しか保存されていない場合は、処理エラーとして、このステップ132の極値編集処理を再実行するようにしてもよい。その場合は、各検査における基準値を緩和して再実行するようにしてもよい。

【0140】サウンド合成に際しては、以上のように編集処理済みの極値リストを利用して、発音時間を長くする制御を行なう。図24のステップ133、134、135に示された合成アルゴリズムにおいて、ステップ133、134では発音時間を長くするためのアルゴリズム、ステップ135では発音時間を短くするためのアルゴリズムを行なう。まず、発音時間を長くするためのアルゴリズムについて説明する。

【0141】ステップ133では、極値リストを参照して、ルーピングのためにつなぎ波形として使用するセグメントに対応する波形データを波形メモリから取り出す。このセグメントは、2つの最大値又は2つの最小値

10

20

30

40

50

の間の波形データである。記録したオリジナルサウンドのどの部分からルーピング用セグメント波形を取り出すべきかは、極値リストが用意されているが故に、全く任意に選択できる。この所望のセグメント波形の選択は、サウンド合成プログラム内に任意にプログラムしておくことによってできるし、ユーザーがマニュアル操作によって任意に選択するようにもできる。例えば、発生しようとする音の性質によって、音の中間部分に対応する波形をループさせるのが好ましい場合や、音の終わりの方の部分の波形をループさせた方が好ましい場合がある。それに限らず、どの部分をループさせるかはユーザーの好みもあるであろうし、サウンド合成プログラムを作成するものの好みもあるであろう。一般的に言って、繰返しは音を単調にするので、サウンドの余り重要でない（そのサウンドをそれほど特徴づけていない）部分のセグメントをループ用のセグメントとして取り出すのがよいであろう。勿論、それに限らず、サウンドを特徴づける部分のセグメントをループ用のセグメントとして取り出すようにしてもよい。なお、ルーピングのために取り出されるセグメント波形データは、SMSデータの全ての種類、つまり周波数トラジェクトリとマグニチュードトラジェクトリ及びストカスティック波形データである。

【0142】ステップ134では、上記のように取り出したセグメント波形を、合成すべきサウンド波形に挿入するための処理を行なう。例えば、オリジナルサウンド波形におけるルーピングを開始するまでの望みの波形

（例えばアタック部の波形、又はアタック部とそれに続く適当な部分の波形）のSMSデータをデータメモリ100から取り出し、これを新しい波形データファイルとしてデータメモリ100の別の記憶位置若しくはその他の適宜のメモリに書き込む。そして、書き込まれた先行波形データに続いて、上記のように取り出したセグメント波形のSMSデータを所望回数だけ繰返し書き込む。セグメント波形を挿入若しくは繰返すときに、データのスムーズな接続が行なわれるように、適当なスムーズ化演算を施すものとする。このスムーズ化演算は、例えば接続部分での補間演算であってもよいし、あるいは、先行する波形の終わりのデータと後続する波形の先頭のデータの値が一致するようにする演算であってもよい。SMSデータにおいてスムーズ化演算の対象とするのは確定的成分のデータであり、ストカスティック成分のデータはスムーズ化演算不要である。延長したい望みの時間分だけ、セグメント波形を繰返し挿入した後は、オリジナル波形の残りのSMSデータを最後の部分として挿入し、メモリに書き込む。この場合も、上記スムーズ化演算を施して、先行するデータと後続するデータの接続がスムーズになされるようにする。

【0143】上述したステップ134の挿入処理は、サウンド発生に関して非実時間的に行なうようにしてい

る。すなわち、発音時間を望みの分だけ延長した波形を作成し、この波形データを新しい波形データファイルとしてデータメモリ100の新たな記憶位置又はその他適宜のメモリに書き込むようにしている。このようにした場合、サウンドを再生発音するときに、メモリからの波形データの順次読み出しを1回だけ行なうことで、延長した発音時間を持つサウンドを合成できる。しかし、これに限らず、シンセサイザー等におけるルーピング処理として知られているような手法で、上述したステップ134の挿入処理と同様な処理を、サウンド発生時に実時間的に行なうようにしてもよい。その場合は、セグメント波形を繰返し書き込む処理は不要であり、ルーピングすべきセグメント波形を指示するデータをステップ133の処理から受け取り、オリジナルサウンド波形を記憶したデータベースの中からこのセグメント波形のデータを繰返し読み出すようにすればよい。変形例としては、発音時間延長のために追加的に繰返されるセグメント波形は、単一のセグメントに限らず、複数セグメントであってもよい。また、1セグメントがビブラートの複数サイクルに対応していてもよい。

【0144】次に、発音時間を短くするためのアルゴリズムについて説明する。発音時間を短くするためのアルゴリズムは、サウンドのいくつかのセグメントを取り除くことを基にしているものである。そのためにステップ135の短縮処理において実行されるアルゴリズムは、周波数トラジェクトリにおける2つのローカル最大値のペアまたは2つのローカル最小値のペアの時間間隔を夫々調べ、取り除きたい時間に適したペアを見つけ出すことからなっている。そのために、周波数トラジェクトリにおけるローカル最大値とローカル最小値のリストを作成し、このリストを参照して、所望の取り除きたい時間に適した極値ペアを見つけ出すようにしてよい。このリストとしては、前述の7ポイント平均値ファイルに基づいて作成した極値リストを用いてよく、その場合、この極値リストは、ステップ131による編集処理を施す前のものであってもよいし、又は施した後のものであってもよい。

【0145】具体的には、音の中央当たりから時間進行方向に沿って極値リストのサーチを開始し、所望の取り除きたい時間に適した2つのローカル最大値のペアまたは2つのローカル最小値のペアを探し出す。こうして、取り除きたい時間に最適の極値ペアを選択する。もし、最大の時間間隔を持つ極値ペアの時間間隔が、所望の取り除きたい時間よりも短い場合は、その最大の時間間隔を持つ極値ペアを、取り除くべき極値ペアとして選択する。次に、図26に示すように、取り除くべきことが選択された極値ペアの間にあるSMSデータのトラジェクトリ部分Bを、オリジナルのSMSデータトラジェクトリA、B、C…から削除する処理を行なう。すなわち、取り除くべきことが選択された極値ペアのうちの最初の

極値よりも前にあるSMSデータトラジェクトリ部分Aをデータメモリ100から取り出して、これを新しい波形データファイルとしてデータメモリ100の新たな記憶位置又はその他適宜のメモリに書き込む。それから、取り除くべきことが選択された極値ペアのうちの2番目の極値よりも後にあるSMSデータトラジェクトリ部分Cをデータメモリ100から取り出して、これを新しい波形データファイルにおけるトラジェクトリ部分Aの次に書き込む。SMSデータトラジェクトリ部分AとCの接続に際しては、前述と同様のスムーズ化演算を行なうものとする。こうして、図27に示すように、トラジェクトリ部分Bを除いた新しいSMSデータファイルが作成される。勿論、削除はSMSデータの全て（周波数、マグニチュード、位相、ストカスティック成分）について行なう。また、波形を短縮すべき時間はユーザーによって任意に選択可能にしてよい。

【0146】上述したステップ135の短縮処理は、サウンド発生に関して非実時間的に行なうようにしている。すなわち、発音時間を望みの分だけ短縮した波形を作成し、この波形データを新しい波形データファイルとしてデータメモリ100の新たな記憶位置又はその他適宜のメモリに書き込むようにしている。しかし、これに限らず、上述したステップ135の短縮処理と同様な処理を、サウンド発生時に実時間的に行なうようにしてもよい。その場合は、取り除くべきセグメントは予めサーチしておき、発音時において、トラジェクトリ部分Aの読み出しが終了した後、取り除くべきセグメントに対応するトラジェクトリ部分Bの読み出しを行なわずに、トラジェクトリ部分Cにジャンプして読み出しを行なうようにすればよい。その場合も、トラジェクトリ部分Aの終わりとトラジェクトリ部分Cの始まりのデータのつながりをスムーズにするための演算処理を行なうのがよい。

【0147】以上述べた例では、発音時間延長用又は短縮用の波形セグメントのサーチは、周波数トラジェクトリにおける極値（すなわちビブラート）を利用して行なっているが、これに限らず、マグニチュードトラジェクトリにおける極値を利用して行なうようにしてもよい。また、発音時間延長用又は短縮用の波形セグメントを見つけ出すための指標としては、極値に限らず、他のものを使用してもよい。この時間変更制御もまた、他の制御と同様に、SMS技術に限らず、他の類似の部分音加算合成技術においても適用可能である。

【0148】ーピッチ分析及び合成ーオリジナルのSMSデータからそのピッチを分析することは、任意の可変ピッチでサウンド合成が行なえるようにするために、極めて重要である。すなわち、オリジナルのSMSデータのピッチが判明していれば、所望の再生ピッチを指定し、該所望再生ピッチとオリジナルピッチとの比に応じてオリジナルのSMSデータの各周波数

データを制御することにより、これらのSMS周波数データを該所望の再生ピッチに対応するものに変更することができる。こうして、変更されたSMSデータは、オリジナルSMSデータの特徴を持つサウンドをそっくり再生できるものでありながら、そのピッチだけがオリジナルとは異なる任意の希望のピッチを持つものとなる。従って、このことを実現可能にするピッチ分析及び合成アルゴリズムは、SMS技術を用いた音楽シンセサイザーにとって極めて重要である。以下、ピッチ分析及び合成アルゴリズムの具体例について説明する。ピッチ分析アルゴリズムはSMSデータ処理器30（図2）で実行され、ピッチ合成アルゴリズムは再生処理器50（図4）で実行される。

【0149】ピッチ分析アルゴリズム：図28はピッチ分析アルゴリズムの一例を示すものである。まず、オリジナルSMSデータの周波数トラジェクトリから各フレーム毎のピッチ $Pf(i)$ を下記式に従い求める（ステップ140）。

【0150】  
【数21】

$$Pf(i) = \frac{\sum_{n=0}^{Np-1} an(i) \frac{fn(i)}{n+1}}{\sum_{n=0}^{Np-1} an(i)}$$

【0151】ここで、 $i$ は特定のフレームを示すフレーム番号、 $Np$ はピッチ分析に使用する部分音の数、 $n$ は部分音の次数を示す変数であり、 $n=0, 1, \dots, Np$ である。 $an(i)$ と $fn(i)$ は、フレーム $i$ における確定的成分中の $n$ 番目の部分音の振幅マグニチュード及び周波数である。式21は、低次の $Np$ 個の部分音の周波数 $fn$ を、夫々の周波数順位の逆数 $1/(n+1)$ と振幅マグニチュード $an$ とで重みづけし、それらの加重平均を算出するものである。この加重平均により、ピッチ $Pf$ を比較的精度良く検出することができる。例えば、 $Np=6$ として、低次の6部分音につき上記加重平均を計算すると、良い結果が得られる。しかし、これに限らず、 $Np=3$ 程度であってもよい。なお、単純には、最低周波数の部分音の周波数 $f0(i)$ をそのフレームのピッチ $Pf(i)$ として検出するようにすることも可能である。しかし、そのような単純な方法よりも、上記のように加重平均によってピッチを検出するようにした方がより聴覚に合っている。

【0152】図30は、上記の加重平均演算に従うフレームピッチ $Pf(i)$ の検出状態を模式的に示す図である。横軸周波数に示された数字1は、検出したフレームピッチ $Pf(i)$ の周波数位置、2, 3, 4, ...は、その2倍、3倍、4倍、...の周波数位置であり、これらは正確な整数倍関係にある。図に示された線スペクトルは、オリジナル周波数データ $fn(i)$ の線スペクトル例であ



る。オリジナルサウンドの線スペクトル $f_n(i)$ は、必ずしも正確な整数倍の周波数関係にはなっていない。図では、加重平均によって求めたピッチの周波数位置は、第1部分音の周波数 $f_0(i)$ とはいくぶん相違していることを示している。

【0153】次に、所定のフレーム範囲にわたる各フレーム毎のピッチ $Pf(i)$ の平均を下記式によって計算し、全体平均ピッチ $Pa$ を得る(ステップ141)。Lは該所定のフレーム範囲におけるフレーム数である。この所定のフレーム範囲としては、オリジナルサウンドのピッチが安定する適当な期間を選ぶのがよい。

【0154】

【数22】

$$Pa = \frac{1}{L} \sum_{i=0}^{L-1} Pf(i)$$

【0155】次に、下記式のように、オリジナルSMSデータにおける各フレーム毎の周波数データ $f_n(i)$ を、そのフレームのピッチ $Pf(i)$ に対する比で表わしたデータ $f'n(i)$ に変換する(ステップ142)。

【数23】 $f'n(i) = f_n(i) / Pf(i)$

ここで、 $n=0, 1, 2, \dots, N-1$ である。次に、下記式のように、各フレーム毎のピッチ $Pf(i)$ を全体平均ピッチ $Pa$ に対する比で表わしたデータ $P'f(i)$ に変換する(ステップ143)。

【数24】 $P'f(i) = Pf(i) / Pa$

【0156】上記数23、24の式によるデータ変換処理によって、SMS周波数データの圧縮化を図ることができると共に、後段での変更制御にあたって処理しやすいデータ表現に変換されることになる。こうして、オリジナルSMSデータにおける絶対的な周波数データ $f_n(i)$ が、相対的な周波数データ群つまり部分音ごとの相対周波数トラジェクトリ $f'n(i)$ 及びフレームピッチトラジェクトリ $P'f(i)$ と、1つの全体平均ピッチデータ $Pa$ とに変換される。これらの変換された周波数データ群 $f'n(i)$ 、 $P'f(i)$ 、 $Pa$ が、SMS周波数データとして、データメモリ100に記憶される。

【0157】ピッチ合成アルゴリズム：図29はピッチ合成アルゴリズムの一例を示すものであり、サウンド合成のためにデータメモリ100から読み出した上記変更されたSMS周波数データ群 $f'n(i)$ 、 $P'f(i)$ 、 $Pa$ を入力し、これに関して下記のような処理を行なう。まず、ステップ150では、合成しようとするサウンドのピッチを制御するためのユーザーの操作に応じた処理を行なう。例えば、ユーザーの操作に応じてピッチ制御パラメータ $Cp$ を発生し、このピッチ制御パラメータ $Cp$ によって全体平均ピッチデータ $Pa$ を変更制御する(例えば乗算する)ことによって、再生サウンドの全体ピッチを指定するデータ $Pd$ を生成する。あるいは、再生サウンドの全体ピッチを指定するデータ $Pd$ を、ユ

ーザーの操作に応じて直接的に発生するようにしてもよい。周知のように、ユーザーの操作に応じたピッチ指定ファクタ又はピッチ制御ファクタには、鍵盤等による音階音指定やピッチバンド等の制御ファクタを含んでいてよい。

【0158】次に、ステップ151では、上記のように確定された所望ピッチ $Pd$ を分析された全体平均ピッチ $Pa$ に置換して、下記式のように相対フレームピッチ $P'f(i)$ と演算することにより、上記式24の逆算を行ない、該所望ピッチ $Pd$ に対応して決定される各フレーム毎の新たなピッチ $Pf(i)$ を求める。

【数25】 $Pf(i) = P'f(i) * Pd$

【0159】次に、ステップ152では、上記のように求められた新たなフレームピッチ $Pf(i)$ とそのフレームに関する各部分音の相対周波数データ $f'n(i)$ とを下記式のように夫々演算することにより、上記式23の逆算を行ない、該所望ピッチ $Pd$ に対応して決定される各フレーム毎の各部分音の絶対周波数データ $f_n(i)$ を求める。ここで、 $n=0, 1, 2, \dots, N-1$ である。

【数26】 $f_n(i) = f'n(i) * Pf(i)$

【0160】以上により、ユーザーの所望するピッチ $Pd$ に対応する絶対周波数で表わされた周波数トラジェクトリ $f_n(i)$ が得られる。このピッチ修正された周波数トラジェクトリ $f_n(i)$ を含むSMSデータに基づきSMSサウンド合成器110でサウンド合成を行なうことにより、所望のピッチ制御がなされたサウンドが得られる。この再生サウンドにおける倍音構成は、その倍音構成に何の制御も加えられない限り、例えば図30に示したようなオリジナルサウンドの倍音構成 $f_0(i)$ 、 $f_1(i)$ 、 $f_2(i)$ 、…を忠実に模倣する(自然音特有の微妙な周波数ずれも模倣している)高品質のものである。また、各データを相対値で表現しているため、倍音構成等を変更するための加工操作も比較的容易に行なえるようになっている。

【0161】なお、所望ピッチ $Pd$ に応じた確定的成分の上記制御と同時に、SMSサウンド合成に使用するストカスティックエンベロープを所望ピッチ $Pd$ に応じて周波数方向に圧縮又は伸長する制御を行なってもよい。このピッチ分析及び合成技術もまた、他の制御と同様に、SMS技術に限らず、他の類似の部分音加算合成技術においても適用可能である。

【0162】一位相分析及び合成—SMS技術において確定的成分の位相データは必須ではないが、位相データを考慮したサウンド合成を行なえば、サウンドの品質をより一層良くすることができる。特に、音のサステイン状態において、適正な位相制御を行なうことは、音の品質を上げるので、望ましい。また、位相を考慮しなかったとすると、ピッチの変更や時間伸長などの変換を位相を含んで行なうことが困難であ



る。そこで、ここでは、確定的成分の位相データの新しい分析及び合成アルゴリズムを提案する。SMS分析されたデータにおける位相トラジェクトリを $\phi_n(i)$ で示す。 $i$ はフレーム番号、 $n$ は部分音の次数である。この位相トラジェクトリ $\phi_n(i)$ における位相値 $\phi_n$ は、各部分音 $n$ 毎の初期位相の絶対値である。新しい位相分析アルゴリズムにおいては、下記式に示すように、この位相値 $\phi_n$ を第1部分音つまり基本成分に対する相対値 $\theta_n(i)$ で表現する。この演算は、SMSデータ処理部30で行なわれる。

【0163】

【数27】

$$\theta_n(i) = \frac{\phi_n(i)}{f_n(i)/f_0(i)} - \phi_0(i)$$

【0164】すなわち、或る部分音の相対位相値 $\theta_n(i)$ は、その絶対位相値 $\phi_n(i)$ を、第1部分音周波数 $f_0(i)$ に対するその部分音周波数 $f_n(i)$ の比で割って\*

$$\phi'_n(i) = [f_n(i)/f_0(i)] * [\theta_n(i) + \phi'_0(i)]$$

【0167】基本的には、上記数式28は、数式27の逆算式である。ただし、 $\phi'_0(i)$ は第1部分音の絶対位相値に相当し、ユーザー操作又は適宜の再生プログラムによって制御可能である。例えば、 $\phi'_0(i) = \phi_0(i)$ とすれば、得られる位相トラジェクトリ $\phi'_n(i)$ はオリジナルの位相トラジェクトリ $\phi_n(i)$ と同じとなる。 $\phi'_0(i) = 0$ とすれば、合成されるサウンドにおける基本成分(第1部分音)の初期位相が0となる。

【0168】この位相トラジェクトリ $\phi'_n(i)$ は、SMSサウンド合成器110において、SMSデータの確定的成分を正弦波合成するとき、各部分音に対応する正弦波波形の初期位相を設定するために利用される。例えば、 $n$ の各値( $n=0, 1, 2, \dots, N-1$ )に対応する正弦波波形を、

$$\sin[2\pi f_n(i)t + \phi'_n(i)]$$

と表現し、これらを加算合計してサウンドを合成するようにしてよい。

【0169】なお、正確な位相の再合成計算は、3次多項式を各部分音の各サンプル毎に計算する必要がある。しかし、そのような計算は、時間がかかり、面倒であるという問題がある。そこで、これを簡略化し、時間のかからない方法で比較的正確な位相の再合成計算を行なえるようにした手法を次に提案する。それは、位相トラジェクトリを使用して周波数トラジェクトリを修正する一種の補間演算からなる。フレームの始まりの周波数を $f_s$ 、終わりの周波数を $f_e$ とし、フレームの始まりの位相を $\phi_s$ 、終わりの位相を $\phi_e$ とする。ここで、もし、単純に、周波数を直線補間したならば、フレームの最後での位相 $\phi_i$ は、次のように表わせる。

【0170】

$$\phi_i = [(f_s + f_e) / 2] * \Delta t + \phi_s$$

\*たものから、第1部分音の絶対位相値 $\phi_0(i)$ を引いたものである。すなわち、高次の部分音ほどその位相の重要度が少なくなるため、それに応じた重み付けを行なってから、第1部分音の位相に対する相対値で表現するようにしている。こうして、位相トラジェクトリ $\phi_n(i)$ は、より小さな値からなる相対位相トラジェクトリ $\theta_n(i)$ に変換され、この状態でデータメモリ100に記憶される。従って、データ圧縮がなされた状態で位相データを記憶することができる。また、第1部分音の相対位相 $\theta_0(i)$ は常に0であるから、これは特に記憶しておく必要がない。

【0165】上記相対位相トラジェクトリ $\theta_n(i)$ に基づき絶対位相トラジェクトリ $\phi_n(i)$ を再合成する処理は、下記式に従って行なう。この演算は再生処理部50で行なわれる。

【0166】

【数28】

ここで、 $\Delta t$ は合成フレームの時間サイズである。 $(f_s + f_e) / 2$ は、始まりの周波数 $f_s$ と、終わりの周波数 $f_e$ の単純平均であり、これに $\Delta t$ を掛けたものは、 $\Delta t$ における周期数を示し、位相に対応している。すなわち、時間 $\Delta t$ からなる1フレームにおいて進行した総位相量に対応している。従って、 $\phi_i$ は単純補間による終わりの位相を示す。次に、 $\phi_e$ と $\phi_i$ の単純平均を次のように求め、これを目標位相 $\phi_t$ とする。

【0171】

$$\phi_t = (\phi_e + \phi_i) / 2$$

この目標位相 $\phi_t$ から、下記式のように目標周波数 $f_t$ を求める。

$$f_t = 2(\phi_t - \phi_s) / \Delta t - f_s$$

ここで、 $\phi_t - \phi_s$ は、目標位相 $\phi_t$ を終わりの位相としたときの時間 $\Delta t$ からなる1フレームにおいて進行する総位相量に対応しており、 $(\phi_t - \phi_s) / \Delta t$ は、そのフレームでの周波数に相当する。この周波数が、始まりの周波数 $f_s$ と目標周波数 $f_t$ との単純平均に相当するものとして、 $f_t$ を求めた式が上記式31である。

【0172】以上の手法で各部分音ごとの位相データを考慮して夫々の周波数データを補間演算し、こうして求めた補間修正済み周波数データを使用して正弦波合成を行なえば、望みの位相合成を比較的正確に行なうことができる。この位相分析及び合成技術もまた、他の制御と同様に、SMS技術に限らず、他の類似の部分音加算合成技術においても適用可能である。

【0173】一周波数及びマグニチュードのデトレンド処理—

デトレンド処理の概略は図3のステップ32に関連して前述した通りである。ここでは、この処理についてその一例につき更に詳しく説明する。ここで述べる例では、

周波数トラジェクトリについては各フレーム毎の基本周波数（これは第1部分音の周波数 $f_0(i)$ 若しくは前述のようなピッチ分析によって分析したフレームピッチ $P_f(i)$ のどちらでもよい）を対象にし、マグニチュードトラジェクトリについては各フレーム毎の平均マグニチュード（そのフレームについての確定的全部分音のマグニチュードの平均値）を対象にし、ストカスティックトラジェクトリについては各フレーム毎のストカスティックゲイン（残差スペクトルエンベロープの全体レベルを示すゲインデータ）を対象にして、夫々処理を行なう。これらの処理対象を以下では要素と呼ぶ。

【0174】まず、サウンドの安定状態に関して、下記式によって、各要素についての時間的変化傾向を示すスロープ $b$ を夫々計算し、各要素についてその変化傾向を見つけ出す。

$$【数32】 b = (y_e - y_0) / (x_e - x_0)$$

ここで、 $y$ はこの式によってその時間的変化傾向を分析しようとする要素の値を示し、 $y_0$ は安定状態の始まりでの要素の値、 $y_e$ は安定状態の終わりでの要素の値である。 $x$ はフレーム番号（つまり時間）を示し、 $x_0$ は安定状態の始まりのフレーム番号、 $x_e$ は安定状態の終わりのフレーム番号である。明らかなように、スロープ $b$ は、変化傾向を示す1次関数の傾き係数に相当する。

【0175】次に、上記スロープ $b$ から、安定状態における各フレーム $x_0, x_1, x_2, \dots, x_e$ に対応してフレーム単位のデトレンド値 $d_i$ を下記式により計算する。

$$【数33】 d_i = (x_i - x_0) * b$$

ここで、 $x_i$ は現在フレーム番号であり、 $i = 0, 1, 2, \dots, e$ についての変数である。

【0176】こうして求めたフレーム単位のデトレンド値 $d_i$ を各要素に対応するSMSデータから引算することにより、デトレンド処理を施す。つまり、スロープ $b$ による癖を取り除いた平坦化されたSMSデータが得られる（ただし、ビブラートやトレモロあるいはその他の微変動は残されている）。周波数要素についてのデトレンド値 $d_i$ の引算は、次のように行なう。このデトレンド値 $d_i$ は基本周波数を基準にしているものであるから、そのフレームにおける各部分音の番号 $n$ （または正確には第1部分音周波数すなわち基本周波数に対する各部分音周波数の比でもよい）をデトレンド値 $d_i$ に掛けたもの $n * d_i$ （ここで $n = 1, 2, \dots, N$ ）を夫々求め、これを対応する部分音周波数から引算する。マグニチュード要素についてのデトレンド値 $d_i$ の引算は、そのフレームにおける各部分音のマグニチュード値からその値 $d_i$ を夫々引算する。ストカスティックゲインについてのデトレンド値 $d_i$ の引算は、そのフレームにおけるストカスティックゲインの値からその値 $d_i$ を引算する。

【0177】デトレンド処理済みのSMSデータは、そのままデータメモリ100に記憶し、サウンド合成のた

めにこれが読み出されるようになっていてよい。通常は、デトレンドしたSMSデータからサウンドを合成する際に、オリジナルのトレンドを再合成して付与する必要はない。すなわち、デトレンドしたままでサウンドを合成してよい。しかし、オリジナルのトレンドをそっくり具備するサウンドを合成したい場合は、トレンド再合成を適宜行なってよい。あるいは、デトレンド処理済みのSMSデータを前述したフォルマント分析やビブラート分析等の各種分析処理の対象として使用するようにしてもよい。

【0178】このデトレンド処理は、SMS分析及び合成にとって必須ではなく、適宜省略できる。しかし、発音時間延長のためにルーピング処理を行なうような場合、SMSデータにデトレンド処理を施しておくことは不自然さのないルーピング（セグメント波形の繰返し）を実現するので、有効である。すなわち、ルーピング用のセグメント波形のSMSデータを作成する目的のみ、補助的にこのデトレンド処理を行なうようにしてもよい。このデトレンド処理技術もまた、SMS技術に限らず、他のサウンド合成技術においても適用可能である。

【0179】—シンギング・シンセサイザのための改良—

この実施例で説明しているシンセサイザは、既に述べた、フォルマントの分析及び合成（制御を含む）技術や、ビブラートの分析及び合成（制御を含む）技術、あるいはノートの転移の際にデータ再生／合成ステップにおいて行なう各種データの補間技術など、その他色々な点で、人声音やボーカルフレーズの合成に適しているものである。以下では、シンギング・シンセサイザとしての応用のために工夫した更なる改良点について説明する。以下で述べる改良点は、SMS分析器20（図2）にて行なうSMS分析処理に関するものである。

【0180】ピッチに同期した分析：SMS技術を使用したシンギング・シンセサイザの特徴の1つは、外部からオリジナルサウンドとして実際のシンギングボイス（人の歌声）を入力し、これをSMS分析することにより、SMSデータを作成し、このSMSデータを自由に加工してからSMS合成を行なうことにより、制御性に富んだ自由なシンギングボイスの合成が行なえることである。ここでは、オリジナルサウンドとして実際のシンギングボイスを入力した場合に有効な、SMS分析の改良を提案する。

【0181】シンギングボイスの特徴の1つは、そのピッチがすばやくかつ連続的に変化することである。そのような場合に分析の精度を上げるために、SMS分析の時間フレームサイズを入力オリジナルサウンドの現在ピッチに従って変化させるようにするとよい。なお、フレームレートは変化させないものとする。フレームサイズを変化させることは、1回のSMS分析のために取り込

む信号の時間長を変えることを意味する。そのために、次のようなステップでSMS分析の或る部分の処理を行なう。この或る部分の処理とは、例えばストカスティック分析のための処理である。

【0182】第1ステップ：過去のフレームの分析結果から入力オリジナルサウンドの基本周波数を得る。

第2ステップ：最後のフレームの基本周波数に応じて現在のフレームサイズを設定する（例えば、周期の4倍の時間とする）。

第3ステップ：時間領域の引算によって残差信号を得る。

第4ステップ：時間領域の残差信号からストカスティック分析を行なう。

【0183】まず、第1ステップであるが、これはSMS分析においては容易に求まる。例えば、これは第1部分音の周波数 $f_0(i)$ 若しくは前述のようなピッチ分析によって分析したフレームピッチ $Pf(i)$ のどちらを基本周波数として用いてもよい。第2ステップのためは、各フレーム毎に異なるフレームサイズを設定することができるようフレキシブルな分析バッファを用意する。こうして用意した各フレームサイズを使用して第3及び第4ステップのストカスティック分析を行なう。第3ステップでは、確定的成分の信号を再生し、これをオリジナル信号から引算して残差信号を得る。第4ステップでは、この残差信号からストカスティック成分のデータを求める。このようなストカスティック分析は、ストカスティック分析用のフレームサイズを、確定的成分分析用のフレームサイズとは異ならせることができるのでよい。例えば、ストカスティック分析用のフレームサイズを、確定的成分分析用のフレームサイズよりも小さくすると、ストカスティック分析結果の時間分解能が良くなり、するどい立上りにおける時間分解能がより良くなる。

【0184】プリエンファシス処理：SMS分析の精度を上げるために、SMS分析を行なう前に、入力音声信号に対してプリエンファシス処理を施すとよい。それから、SMS分析の最後に、プリエンファシスに対応したデエンファシス処理を施す。このようなプリエンファシス処理は、より高い周波数の部分音まで分析できるようにするので、好ましい。

【0185】残差信号に対するハイパスフィルタ処理：通常、シンギングボイスのストカスティック成分は高周波数である。200Hz以下のストカスティック信号は極めて少ない。従って、SMS分析において、SMS分析された確定的成分信号をオリジナルサウンド信号から引算することによって求めた残差信号に基づきストカスティック分析を行なう前に、この残差信号に対してハイパスフィルタ処理を施すのがよい。それとは別に、オリジナルサウンド信号からの確定的成分信号の引算は、音声によくみられる早いピッチ変化のために低周波数にお

いていくつかの問題を持っている。そのためにハイパスフィルタを使用するとよい。例えばハイパスのカットオフ周波数を800Hz程度に設定するとよい。このフィルタリングが実際のストカスティック信号を差し引かないようにするための妥協策は、分析しようとするサウンドの部分に従属してそのカットオフ周波数を変化させることである。例えば、多くの確定的成分を持つが、ストカスティック成分は僅かしか持たないサウンドの部分においては、カットオフ周波数をより高くすることができる。その逆に、多くのストカスティック成分を持つサウンドの部分においては、カットオフ周波数をより低くしなければならない。

【0186】ーボークアルフレーズ合成の具体例ー

以上に述べたこの発明のシンセサイザを使用してボークアルフレーズを合成するためには、まず、複数の音素 (phoneme) 及び重なり音 (diphone) のデータベースを作成する。そのため、各音素及び重なり音のサウンドを入力してSMS分析を夫々行ない、それらのSMSデータを作成し、データメモリ100に夫々記憶することによりそれらのデータベースを作成する。こうして、作成されたデータベースから、ユーザーの制御に基づき、所望のボークアルフレーズを構成するために必要な複数の音素及び／又は重なり音のSMSデータを読み出し、これらを時系列的に組合せて該ボークアルフレーズに対応するSMSデータを作成する。作成された所望のボークアルフレーズに対応するSMSデータの組合せは、メモリに記憶しておき、望みのときにこれを読み出すことによりボークアルフレーズのサウンド合成を行なうようにしてよい。あるいは、作成された所望のボークアルフレーズに対応するSMSデータの組合せに対応するサウンドを実時間的にSMS合成することにより、該ボークアルフレーズのサウンド合成を行なうようにしてもよい。

【0187】入力サウンドの分析にあたっては、例えば、入力サウンドが、単一の音素又は重なり音であるとみなしてSMS分析を行なうようにしてよい。単一の音素又は重なり音における周波数成分は、そのサウンドの安定状態においては、あまり変化しないので、分析がしやすい。従って、例えば、望みの或る音素を分析しようとする場合、サウンドの安定状態においてその音素の特徴が現われるサウンドを入力してやればよい。このような音素又は重なり音の分析つまり人声音の分析のために、従来知られたSMS分析を行なうのみならず、この明細書で説明した様々な改良（フォルマント分析やビブラート分析など、その他）を併せて行なうことは、人声音の分析及びその自由な可変的合成にとって、きわめて有益である。

【0188】ーSMSデータの対数表現ー

SMSデータにおける周波数データは、従来はHzまたはラジアンに対応するリニア表現からなるものであった。しかし、これに限らず、この周波数データを対数表

現で表わしてもよい。そうすると、前述した様々な演算、例えばピッチ変更のための演算など、における周波数データの乗算を、簡単な加算によって置き換えることができる。

【0189】一ストカスティックエンベロープのスムーズ化—

与えられたサウンドのストカスティック表現データを計算するための方法の1つは、残差スペクトルエンベロープについてのラインセグメント近似によるものである。ストカスティックデータの周波数エンベロープを一旦計算したら、このエンベロープをローパスフィルタで処理してスムーズにするのが良い。この処理によって、合成されるノイズ信号はスムーズなものとなる。

【0190】—デジタルウェーブガイド技術への応用—デジタルウェーブガイド理論に従って音を合成する技術が知られている（例えば米国特許第4,984,276号）。これを極めてシンプルに示すと図31のようであり、閉鎖されたウェーブガイドネットワーク160に、励振関数発生器161から発生した励振関数信号を入力し、ウェーブガイドネットワーク160において設定されているパラメータに従って信号処理を行なうことにより、該パラメータによって設定した所望音色の出力サウンドを得るものである。このようなデジタルウェーブガイド理論に従う楽音合成技術にSMS技術を応用することを考えると、励振関数発生器161をSMSサウンド合成システムによって構成し、SMS合成したサウンド信号をウェーブガイドネットワーク160に対する励振関数信号として使用する方法が考えられる。

【0191】更に具体的に考察すると、一例として、図32のような処理手順で、ウェーブガイドネットワーク160に対する励振関数信号をSMS合成する方法が考えられる。まず、ウェーブガイドネットワーク160から出力したい望みのサウンドに対応するオリジナルサウンドの信号を、ウェーブガイドネットワーク160で設定するフィルタ特性とは反対の特性に設定した逆フィルタ回路で処理する（ステップ162）。この出力が望みの励振関数信号に対応する。次に、この望みの励振関数信号をSMS分析器で分析し（ステップ163）、これに対応するSMSデータを得る。このSMSデータをデータメモリに適宜記憶しておく。それから、このSMSデータをデータメモリから読み出し、ユーザーの制御に応じて適宜変更を施し（ステップ164）、SMS合成器でサウンド合成する（ステップ165）。こうして合成されたサウンド信号を励振関数信号としてウェーブガイドネットワーク160に入力する。

【0192】このような方法の利点は、ウェーブガイドネットワーク160のパラメータを変えずに、SMS合成による励振関数信号を変えることにより、望みのサウンドを合成できる点であり、ウェーブガイドネットワークのパラメータ解析を簡単化することができる。すなわ

ち、サウンドを合成する際の所望の可変制御は、SMSデータの変更制御によってかなり実現できることになり、その分、ウェーブガイドネットワーク側での可変制御のためのパラメータ解析を簡単化することができる、ということが期待できる点である。

【0193】以上の実施例に基づき抽出されるこの出願の発明若しくは実施態様のいくつかを要約して列挙すると次のようである。

1. オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第2のステップと、抽出されたサウンドパラメータに対応する特徴を前記分析データから取り除く第3のステップと、前記特徴が取り除かれた分析データに対して、サウンドパラメータに対応する特徴を付加する第4のステップと、この特徴が付加された分析データに基づき、サウンド波形を合成する第5のステップとを備えたサウンドを分析し合成するための方法。
2. 前記第4のステップは、前記サウンドパラメータを変更するためのステップを含み、変更されたサウンドパラメータに対応する特徴を前記分析データに付加することを特徴とする前記1項の方法。
3. 前記特徴が取り除かれた分析データと、前記サウンドパラメータとをメモリに記憶するステップを更に具えることを特徴とする前記1項の方法。
4. 前記サウンドパラメータは、前記分析データとは異なるデータ表現で表現されたものであることを特徴とする前記1項の方法。
5. 前記第4のステップは、前記サウンドパラメータから前記分析データのデータ表現に対応するデータ表現からなる付加データを再生するステップを含み、この付加データを前記分析データに付加することを特徴とする前記1項の方法。
- 【0194】6. 前記第4のステップの前に、少なくとも2つの異なるサウンド又はサウンド部分に関する前記分析データを補間すると共に、該異なるサウンド又はサウンド部分に関する前記サウンドパラメータを補間するステップを更に含み、前記第4のステップでは、補間された前記分析データに対して、補間された前記サウンドパラメータに対応する特徴を付加することを特徴とする前記1項の方法。
7. 前記分析データは、オリジナルサウンド波形を構成する部分音の周波数及びマグニチュードを示すデータを含むことを特徴とする前記1項の方法。
8. 前記分析データは、オリジナルサウンド波形を構成する部分音の周波数及びマグニチュードを示す確定的成分のデータと、前記オリジナルサウンド波形の前記確定

的成分に対する残差成分に対応するストカスティックデータとを含むことを特徴とする前記1項の方法。

9. 前記第1のステップでは、前記オリジナルサウンドを異なる時間フレームで分析することによって得られる各時間フレーム毎の分析データを提供し、前記第2のステップでは、各時間フレーム毎の分析データに基づき、各時間フレーム毎に前記サウンドパラメータを抽出することを特徴とする前記1項の方法。

10. 前記第1のステップでは、前記オリジナルサウンドを異なる時間フレームで分析することによって得られる各時間フレーム毎の分析データを提供し、前記第2のステップでは、各時間フレーム毎の分析データに基づき、複数の時間フレームに共通の前記サウンドパラメータを抽出することを特徴とする前記1項の方法。

11. 前記サウンドパラメータに対応する特徴は周波数成分に関するものであり、前記第3のステップにおける分析データからの該特徴の取り除きは、分析データにおける周波数データを変更することからなることを特徴とする前記1項の方法。

12. 前記サウンドパラメータに対応する特徴はマグニチュード成分に関するものであり、前記第3のステップにおける分析データからの該特徴の取り除きは、分析データにおけるマグニチュードデータを変更することからなることを特徴とする前記1項の方法。

【0195】13. オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナルサウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第2のステップと、抽出されたサウンドパラメータに対応する特徴を前記分析データから取り除く第3のステップとを備え、前記特徴が取り除かれた分析データと、前記サウンドパラメータとの組合せによって前記オリジナルサウンド波形を表現することを特徴とするサウンドを分析するための方法。

14. 前記特徴が取り除かれた分析データと、前記サウンドパラメータとをメモリに記憶するステップを更に具えることを特徴とする前記13項の方法。

15. 前記分析データは、オリジナルサウンド波形を構成する部分音の周波数及びマグニチュードを示す確定的成分のデータと、前記オリジナルサウンド波形の前記確定的成分に対する残差成分に対応するストカスティックデータとを含むことを特徴とする前記13項の方法。

【0196】16. オリジナルサウンドを分析することにより、該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する第1のステップと、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータを、前記オリジナル

サウンドにおける前記要素についての固有の特性を示すサウンドパラメータとして抽出する第2のステップと、前記サウンドパラメータを変更するための第3のステップと、前記分析データに対して、前記サウンドパラメータに対応する特徴を付加する第4のステップと、この特徴が付加された分析データに基づき、サウンド波形を合成する第5のステップとを備えたサウンドを分析し合成するための方法。

17. 前記分析データは、オリジナルサウンド波形を構成する部分音の周波数及びマグニチュードを示す確定的成分のデータと、前記オリジナルサウンド波形の前記確定的成分に対する残差成分に対応するストカスティックデータとを含むことを特徴とする前記16項の方法。

【0197】18. オリジナルサウンドの分析に基づき該オリジナルサウンド波形を構成する複数の成分を示す分析データを提供する分析手段と、前記分析データから所定のサウンド要素に関する特徴を分析して、分析した該特徴を示すデータをサウンドパラメータとして抽出すると共に、抽出されたサウンドパラメータに対応する特徴を前記分析データから取り除くデータ処理手段と、前記特徴が取り除かれた分析データと前記サウンドパラメータとを記憶する記憶手段と、前記分析データとサウンドパラメータを前記記憶手段から読み出し、読み出した分析データに対して該サウンドパラメータに対応する特徴を付加するデータ再生手段と、データ再生手段で再生された分析データに基づき、サウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

19. 前記サウンドパラメータを変更するための変更手段を更に具え、前記データ再生手段では変更されたサウンドパラメータに対応する特徴を前記分析データに対して付加し、これにより、合成するサウンドを制御することができることを特徴とする前記18項のサウンド波形合成装置。

20. 前記変更手段は、ユーザーの操作に応じて前記サウンドパラメータを変更できるものであることを特徴とする前記19項のサウンド波形合成装置。

21. 前記データ再生手段は、少なくとも2つの異なるサウンド又はサウンド部分に関する前記分析データを補間すると共に、該異なるサウンド又はサウンド部分に関する前記サウンドパラメータを補間する補間手段を含み、補間された分析データに対して補間されたサウンドパラメータに対応する特徴を付加することを特徴とする前記18項のサウンド波形合成装置。

22. 前記分析データは、オリジナルサウンド波形を構成する部分音の周波数及びマグニチュードを示す確定的成分のデータと、前記オリジナルサウンドの前記確定的成分に対する残差成分に対応するストカスティックデータとを含むことを特徴とする前記18項のサウンド波形合成装置。

【0198】23. 部分音を示すデータを含む波形分析

データと、オリジナルサウンドから抽出された所定のサウンド要素に関する特徴を示すサウンドパラメータとを記憶している記憶手段と、前記波形分析データとサウンドパラメータを前記記憶手段から読み出す読み出し手段と、読み出されたサウンドパラメータを変更するための制御を行なう制御手段と、前記読み出した波形分析データを、前記制御されたサウンドパラメータによって変更するデータ変更手段と、データ変更手段で変更された波形分析データに基づき、サウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

24. 前記記憶手段に記憶される前記波形分析データは、更にスペクトルエンベロープデータを含んでおり、前記サウンド合成手段は、前記波形分析データに含まれる前記部分音を示すデータに基づき各部分音の波形を発生する確定的波形発生手段と、前記波形分析データに含まれるスペクトルエンベロープデータに基づいて定まるスペクトルマグニチュードを持つストカスティックなスペクトル構成からなるストカスティック波形を発生するストカスティック波形発生手段と、前記各部分音の波形とストカスティック波形とを組み合わせることによりサウンド波形を合成する手段とを具えることを特徴とする前記23項のサウンド波形合成装置。

【0199】25. オリジナルサウンドをスペクトル分析したデータを提供する第1の手段と、前記スペクトル分析されたデータからフォルマント構造を検出し、検出したフォルマントを記述するパラメータを生成する第2の手段と、前記スペクトル分析されたデータから前記検出されたフォルマント構造を差引き、残余のスペクトルデータを生成する第3の手段とを備え、前記残余のスペクトルデータと前記パラメータとの組合せによって前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

26. 前記フォルマントを制御するために前記パラメータを可変制御する第4の手段と、前記パラメータに基づきフォルマント構造を再生し、再生されたフォルマント構造を前記残余のスペクトルデータに付加し、制御されたフォルマント構造を有するスペクトルデータを作成する第5の手段とを更に備えたことを特徴とする前記25項のサウンド波形合成装置

27. 前記第5の手段で作成されたスペクトルデータに基づきサウンド波形を合成するサウンド合成手段を更に備えたことを特徴とする前記26項のサウンド波形合成装置。

【0200】28. 前記第1の手段は、前記オリジナルサウンドを異なる時間フレームで分析することによって得られた各時間フレーム毎のスペクトル分析データを提供するものであり、前記第2の手段は、各時間フレーム毎のスペクトル分析データに基づき、各時間フレーム毎にフォルマント構造をそれぞれ検出し、検出したフォルマントを記述するパラメータを生成するものであり、前

記第3の手段は、各時間フレーム毎のスペクトル分析データから前記各時間フレーム毎に検出されたフォルマント構造を差引き、残余のスペクトルデータを各時間フレーム毎に生成することを特徴とする前記25項のサウンド波形合成装置。

29. 前記第2の手段は、前記スペクトル分析データにおける各線スペクトルのマグニチュードに基づき、2つのローカル最小値とそれによって囲まれた1つのローカル最大値とからフォルマントと推定される1又は複数の山を検出する手段と、検出した各山毎に所定の関数近似によりフォルマントエンベロープを近似し、この近似により少なくともフォルマント中心周波数とそのピークレベルを記述するデータを含むフォルマントパラメータを求める手段とを有することを特徴とする前記25項のサウンド波形合成装置。

30. 前記フォルマントエンベロープの近似は、指数関数近似によって行なうことを特徴とする前記29項のサウンド波形合成装置。

31. 前記フォルマントエンベロープの近似は、二等辺三角形関数近似によって行なうことを特徴とする前記29項のサウンド波形合成装置。

【0201】32. オリジナルサウンドの分析によって得た複数のサウンド部分を示す部分音データのセットを提供するものであり、各部分音データは周波数データを含み、前記部分音データのセットを時間関数で提供する第1の手段と、前記部分音データにおける周波数データの時間関数からオリジナルサウンドにおけるビブラートを検出し、検出したビブラートを記述するパラメータを生成する第2の手段と、前記部分音データにおける周波数データの時間関数から前記検出されたビブラートの特徴を取り除き、修正された周波数データの時間関数を生成する第3の手段とを備え、前記修正された周波数データの時間関数を含む前記部分音データと前記パラメータとの組合せによって時間的に変化する前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

33. ビブラートを制御するために前記パラメータを可変制御する第4の手段と、前記パラメータに基づきビブラート関数を発生し、発生されたビブラート関数によって前記修正された周波数データの時間関数にビブラートを付与する第5の手段と、ビブラート付与された周波数データの時間関数を含む前記部分音データに基づきサウンド波形を合成するサウンド合成手段とを更に備えたことを特徴とする前記32項のサウンド波形合成装置。

34. 前記第2の手段は、前記周波数データの時間関数をスペクトル解析することにより、ビブラートを検出するものであり、前記第3の手段は、解析された前記周波数データの時間関数のスペクトルデータから、検出されたビブラートの成分を除去し、その結果たる時間関数のスペクトルデータを逆フーリエ変換することにより、修

10

20

30

40

50

正された周波数データの時間関数を生成することを特徴とする前記32項のサウンド波形合成装置。

35. 前記第2の手段は、所定の1又は複数の低次の部分音の周波数データの時間関数について前記スペクトル解析を行なうことにより、ビブラートを検出することを特徴とする前記34項のサウンド波形合成装置。

【0202】36. オリジナルサウンドの分析によって得た複数のサウンド部分を示す部分音データのセットを提供するものであり、各部分音データはマグニチュードデータを含み、前記部分音データのセットを時間関数で提供第1の手段と、前記部分音データにおけるマグニチュードデータの時間関数からオリジナルサウンドにおけるトレモロを検出し、検出したトレモロを記述するパラメータを生成する第2の手段と、前記部分音データにおけるマグニチュードデータの時間関数から前記検出されたトレモロの特徴を取り除き、修正されたマグニチュードデータの時間関数を生成する第3の手段とを備え、前記修正されたマグニチュードデータの時間関数を含む前記部分音データと前記パラメータとの組合せによって時間的に変化する前記オリジナルサウンド波形を表現することを特徴とするサウンド波形合成装置。

37. トレモロを制御するために前記パラメータを可変制御する第4の手段と、前記パラメータに基づきトレモロ関数を発生し、発生されたトレモロ関数によって前記修正されたマグニチュードデータの時間関数にトレモロを付与する第5の手段と、トレモロ付与されたマグニチュードデータの時間関数を含む前記部分音データに基づきサウンド波形を合成するサウンド合成手段とを更に備えたことを特徴とする前記36項のサウンド波形合成装置。

【0203】38. オリジナルサウンドのスペクトル構造を示すスペクトルデータを提供する第1の手段と、前記スペクトルデータに基づき、そのスペクトルエンベロープに概ね適合しているただ1本のチルトラインを検出し、検出したチルトラインを記述するチルトパラメータを生成する第2の手段と、スペクトルの傾きを制御するために、前記チルトパラメータを可変制御する第3の手段と、制御されたチルトパラメータに基づき前記スペクトルデータのスペクトル構造を制御する第4の手段と、制御されたスペクトルデータに基づきサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

39. 前記第1の手段は、前記オリジナルサウンドを異なる時間フレームで分析することによって得られた各時間フレーム毎のスペクトルデータを提供するものであり、前記第2の手段は、各時間フレーム毎のスペクトルデータに基づき、各時間フレーム毎のチルトラインを検出し、これらのチルトラインを示すデータに基づきこれらの相関を示すただ1つの前記チルトパラメータを生成するものであり、更に、前記各時間フレーム毎のスペク

トルデータを、前記チルトパラメータを使用して正規化する第5の手段を具備し、かつ、前記第4の手段は、制御されたチルトパラメータに基づき前記正規化されたスペクトルデータの正規化を解除することを特徴とする前記38項のサウンド波形合成装置。

【0204】40. オリジナルサウンドを構成する部分音のスペクトルデータを複数の時間フレームに対応して提供する第1の手段と、一連の時間フレームにおける前記部分音スペクトルデータにおける周波数データに基づき前記オリジナルサウンドの平均ピッチを検出し、ピッチデータを生成する第2の手段と、前記ピッチデータを可変制御するための第3の手段と、制御されたピッチデータに応じて前記部分音スペクトルデータにおける周波数データを修正する第4の手段と、前記修正された周波数データを含む前記部分音スペクトルデータに基づき、可変制御されたピッチを持つサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

41. 前記第1の手段は、更に、前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナルサウンドから引いた残りである残差成分波形に対応するストカスティックデータを提供するものであり、前記第4の手段は、更に、制御されたピッチデータに応じて前記ストカスティックデータの周波数特性を制御することを特徴とする前記40項のサウンド波形合成装置。

42. 前記部分音スペクトルデータにおける周波数データを前記検出した平均ピッチを基にした相対値に変換する手段を更に備え、前記第4の手段は、制御されたピッチデータに応じて前記相対値を絶対値に変換し、前記修正された周波数データを得ることを特徴とする前記40項のサウンド波形合成装置。

43. 前記第2の手段は、各時間フレーム毎に、所定の複数の低次の部分音の周波数をそのマグニチュードに応じて重みづけして平均化することによりフレームピッチをそれぞれ求め、各フレームピッチを平均化することにより平均ピッチを検出することを特徴とする前記40項のサウンド波形合成装置。

【0205】44. オリジナルサウンドを構成する部分音のスペクトルデータと、前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナルサウンドから引いた残りである残差成分波形に対応するストカスティックデータと、前記オリジナルサウンドの特定されたピッチを示すピッチデータとを記憶するものであり、前記部分音スペクトルデータにおける各周波数データを、前記ピッチデータが示す特定の周波数を基にした相対値で表わしたデータで記憶している記憶手段と、前記記憶手段に記憶したデータを読み出すための手段と、前記記憶手段から読み出された前記ピッチデータを可変制御するための制御手段と、前記記憶手段から読み出された前記部分音スペクトルデータにおける周波数データの相対値を、前記制御されたピッチデータに応じて絶対値に変



換する演算手段と、変換された周波数データと前記記憶手段から読み出された前記部分音スペクトルデータにおけるマグニチュードデータとに基づき部分音波形を合成し、かつ、前記記憶手段から読み出された前記ストカスティックデータに基づき前記残差成分波形を合成し、前記部分音波形と前記残差成分波形を組み合わせたサウンド波形を合成するサウンド合成手段とを備えたサウンド波形合成装置。

45. 前記記憶手段に記憶する前記部分音スペクトルデータには位相データが含まれており、この位相データは、各部分音の位相を基本の部分音の位相を基準にした相対値で表わされており、更に、前記記憶手段から読み出された前記部分音スペクトルデータにおける位相データの相対値を絶対値に変換する手段を具備し、前記サウンド合成手段では、変換された位相データと前記周波数データ及びマグニチュードデータとに基づき前記部分音波形を合成することを特徴とする前記44項のサウンド波形合成装置。

【0206】46. ウェーブガイドをモデルした閉鎖ネットワークであって、振動関数信号を該閉鎖ネットワーク内に導入し、前記ウェーブガイドにおける信号の遅延と散乱をシミュレートするパラメータによって決定される処理を該信号に対して施すことにより、サウンド信号を合成する閉鎖型ウェーブガイドネットワーク手段と、前記振動関数信号を発生するための振動関数発生手段とを具備し、前記振動関数発生手段は、オリジナル信号波形を構成する部分音のスペクトルデータと、前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナル信号波形から引いた残りである残差成分波形に対応するストカスティックデータとを記憶している記憶手段と、前記記憶手段に記憶したデータを読み出すための手段と、前記記憶手段から読み出されたデータを可変制御するための制御手段と、前記部分音スペクトルデータに基づき部分音波形を合成し、かつ、前記ストカスティックデータに基づき前記残差成分波形を合成し、前記部分音波形と前記残差成分波形を組み合わせた波形信号を合成する波形合成手段とを有しており、合成された波形信号を前記振動関数信号として前記ウェーブガイドネットワークに与えるようにしたことを特徴とするサウンド波形合成装置。

47. 前記記憶手段は、所定の音楽要素に関する特徴を示すパラメータを更に記憶しており、前記制御手段は、前記パラメータを可変制御すると共に、制御されたパラメータによって前記部分音スペクトルデータ及びストカスティックデータを可変制御することを特徴とする前記46項のサウンド波形合成装置

【0207】48. オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブ

ラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、前記データリストを参照して、任意の波形セグメントを選択するステップと、選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から抜き出すステップと、抜き出したスペクトルデータ列を繰り返すことにより前記波形セグメントの繰り返しに対応するスペクトルデータ列を作成するステップと、前記繰り返しに対応するスペクトルデータ列を使用して、延長された長さを持つサウンド波形を合成するステップとを備えたサウンドを分析し合成する方法。

49. 前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナル波形から引いた残りである残差成分波形に対応するストカスティックデータを、複数の時間フレームに対応して順次に提供するステップと、前記選択した波形セグメントに対応する前記ストカスティックデータ列を前記オリジナル波形のストカスティックデータ列から抜き出すステップと、抜き出したストカスティックデータ列を繰り返すことにより前記波形セグメントの繰り返しに対応するストカスティックデータ列を作成するステップと、前記繰り返しに対応するストカスティックデータ列を使用して、延長された長さを持つストカスティック波形を合成し、これを前記サウンド波形に組み込むステップとを更に備えたことを特徴とする前記48項の方法。

【0208】50. オリジナル波形を構成する部分音のスペクトルデータを、複数の時間フレームに対応して順次に提供するステップと、前記複数の時間フレームのスペクトルデータ列から前記オリジナル波形におけるビブラート変動を検出し、この変動の少なくとも1サイクルに対応する長さを持つ1又は複数の波形セグメントを指摘するデータリストを作成するステップと、前記データリストを参照して、任意の波形セグメントを選択するステップと、選択した波形セグメントに対応する前記スペクトルデータ列を前記オリジナル波形のスペクトルデータ列から取り去り、その前後で残された2つのスペクトルデータ列を接続し、短縮されたスペクトルデータ列を作成するステップと、前記短縮されたスペクトルデータ列を使用して、短縮された長さを持つサウンド波形を合成するステップとを備えたサウンドを分析し合成する方法。

51. 前記部分音スペクトルデータに対応する確定的成分波形を前記オリジナル波形から引いた残りである残差成分波形に対応するストカスティックデータを、複数の時間フレームに対応して順次に提供するステップと、前記選択した波形セグメントに対応する前記ストカスティックデータ列を前記オリジナル波形のストカスティックデータ列から取り去り、その前後で残された2つのストカスティックデータ列を接続し、短縮されたストカステ



ックデータ列を作成するステップと、前記短縮されたストカスティックデータ列を使用して短縮された長さを持つストカスティック波形を合成し、これを前記サウンド波形に組み込むステップとを更に備えた前記50項の方法。

【0209】

【発明の効果】 以上の通り、この発明によれば、オリジナルサウンドの分析データから所定のサウンド要素に関する特徴を分析し、分析した該特徴を示すデータをサウンドパラメータとして抽出するようにしたので、例えばフォルマントやビブラートなどのような様々なサウンド要素に関して、オリジナルの特徴を示している品質のよいサウンドパラメータを得ることができる。従って、このパラメータをサウンド波形合成に際して利用すれば、品質のよい各種音楽的特徴の合成を行なうことができる。しかも、サウンドパラメータとして分析データから分離抽出されているため、その可変制御が容易であり、ユーザーによる自由な音楽制御に適したものである。また、抽出されたサウンドパラメータに基づき、前記オリジナルサウンドの分析データから時変動成分を取り除くようにしているため、分析データの構造が簡単化され、データ圧縮が期待できるものである。また、時変動成分が取り除かれた分析データ又は合成されたサウンド波形に対して任意の時間変化を付与することで任意の時間変化特性が付与されたサウンド波形を合成するようにしているので、オリジナルの特徴を示している品質のよいサウンド波形に対して任意の時間変化特性を付与することができ、品質と制御性に優れたサウンド波形合成を行なうことができる。このように、サウンドパラメータを分析データから抽出分離し、時変動成分が取り除かれた分析データとサウンドパラメータの組み合わせによってオリジナルサウンド波形を表現するデータを提供し、これに基づきサウンド波形を合成する技術は、種々の効果が期待できるものである。

【図面の簡単な説明】

【図1】 この発明の一実施例に係る音楽シンセサイザの全体を示すブロック図。

【図2】 図1における分析部の一実施例を示すブロック図。

【図3】 図2におけるSMSデータ処理部の一実施例を示すブロック図。

【図4】 図1における合成部の一実施例を示すブロック図。

【図5】 図4における再生処理部の一実施例を示すブロック図。

【図6】 この発明に従うフォルマント抽出及び操作システムの一実施例を示すブロック図。

【図7】 図6に入力されるSMS分析されたデータのうち1フレーム分の確定的部分のデータすなわち線スペクトルデータの一例を示す線スペクトル図。

【図8】 図6に入力されるSMS分析されたデータのうち1フレーム分のストカスティックエンベロープの一例を示すスペクトルエンベロープ図。

【図9】 図6の実施例に従って、線スペクトルにおけるフォルマントを指数関数近似によって検出する状態を説明するための図。

【図10】 検出したフォルマントの特徴を差し引いて平坦化された線スペクトル構造の一例を示す図。

【図11】 この発明に従うフォルマント抽出及び操作システムの別の実施例を示すブロック図。

【図12】 図11の実施例に従って、線スペクトルにおけるフォルマントを三角形関数近似によって検出する状態を説明するための図。

【図13】 フォルマントの三角形関数近似の第1ステップとして、フォルマントの山を検出する状態を説明するための図。

【図14】 フォルマントの三角形関数近似の第2ステップとして、二等辺三角形近似のために、フォルマント中心周波数の位置で線スペクトルを折り返した状態を模式的に示す図。

【図15】 フォルマントの三角形関数近似の第3ステップとして、二等辺三角形近似が達成できた状態を示す図。

【図16】 検出したフォルマントをトラジェクトリに割当てて様子模式的に示す図。

【図17】 この発明に従うビブラート分析システムの一実施例を示すブロック図。

【図18】 図17の実施例において、周波数トラジェクトリの時間関数をフーリエ変換することにより求められるスペクトルエンベロープの一例を示す図。

【図19】 図18のスペクトルからビブラート成分を取り除いた状態を示すスペクトルエンベロープの一例を示す図。

【図20】 図17の実施例において、図18のようなスペクトル特性からビブラートレートを放物線近似によって計算する一例を拡大して示す図。

【図21】 この発明に従うビブラート合成アルゴリズムの一実施例を示すブロック図。

【図22】 この発明に従うスペクトルチルトの分析及び合成アルゴリズムの一実施例を示すブロック図。

【図23】 図22の実施例に従って、SMS分析されたデータのうち1フレーム分の確定的部分のデータすなわち線スペクトルから分析されるチルトの一例を示す図。

【図24】 この発明に従う発音時間変更アルゴリズムの一実施例を示すブロック図。

【図25】 図24の実施例に従って分析されるビブラート極値とスロープの一例を示す図。

【図26】 図24の実施例における、発音時間短縮のための取り除き部分の分析例を示す図。

【図27】 図25の例において、分析した取り除き部分

65

を波形データから取り除いて発音時間を短縮したデータの例を示す図。

【図28】この発明に従うピッチ分析アルゴリズムの一実施例を示すブロック図。

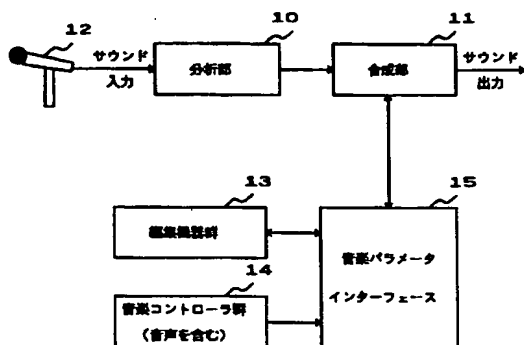
【図29】この発明に従うピッチ合成アルゴリズムの一実施例を示すブロック図。

【図30】図28のピッチ分析アルゴリズムにおける1フレームについてのピッチ検出を説明するためのスペクトル図。

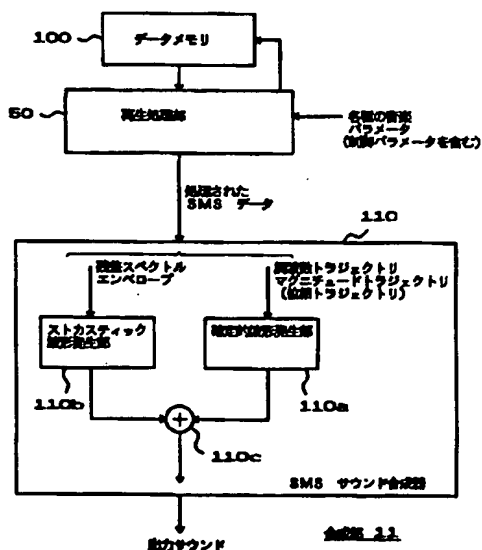
【図31】この発明に従うSMS技術をデジタルウェーブガイド理論による楽音合成技術に適用する一実施例を示すブロック図。

【図32】図31における励振関数発生器に対するSM

【図1】



【図4】



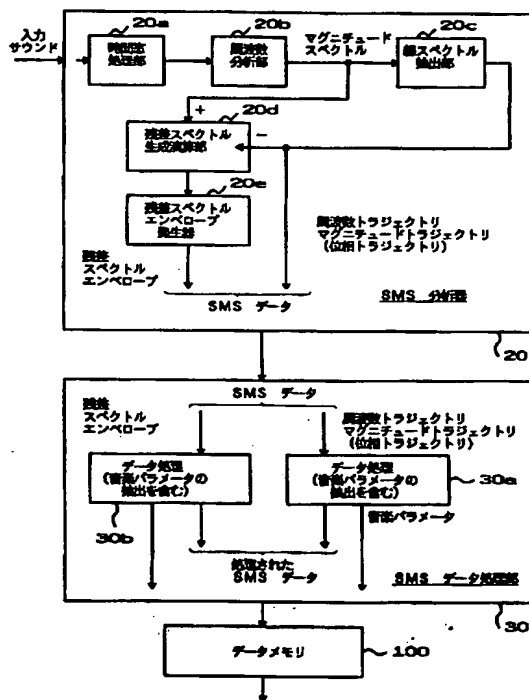
66

S分析及び合成技術の適用例を示すブロック図。

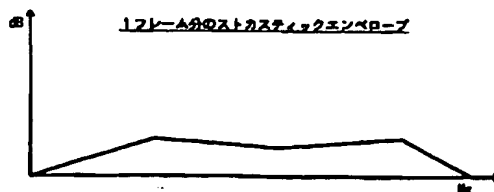
【符号の説明】

- 10 分析部
- 11 合成部
- 13 編集機器群
- 14 音楽コントローラ群
- 15 音楽パラメータインターフェース部
- 20 SMS分析器
- 30 SMSデータ処理部
- 50 再生処理部
- 100 データメモリ
- 110 SMSサウンド合成器

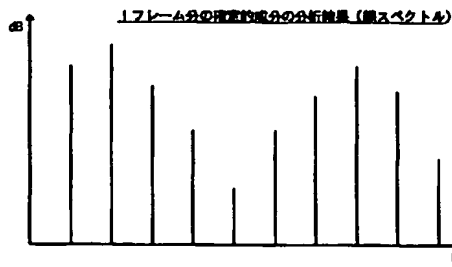
【図2】



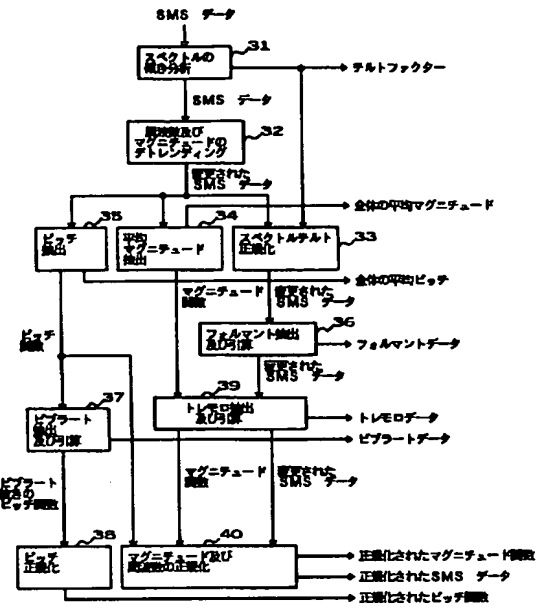
【図8】



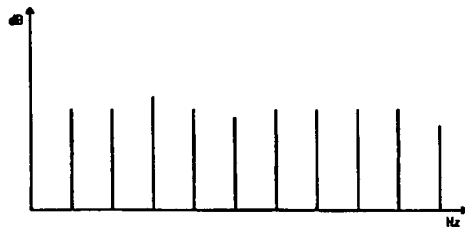
【図7】



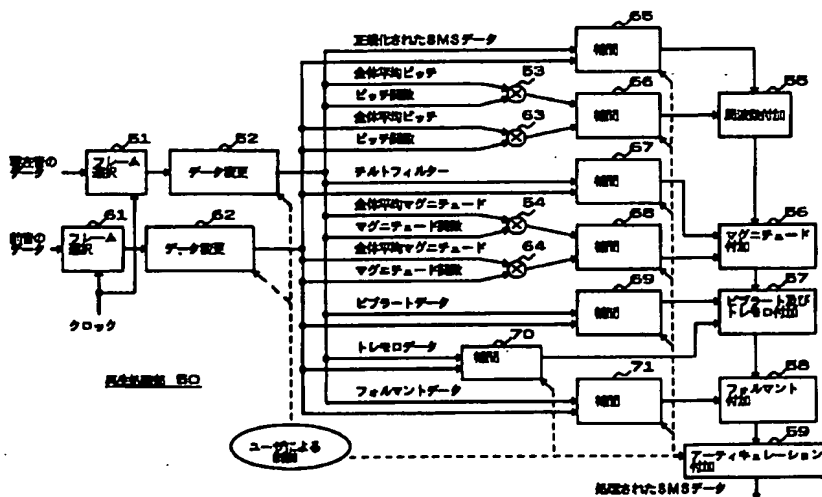
【図3】



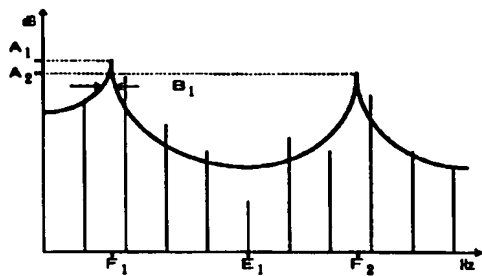
【図10】



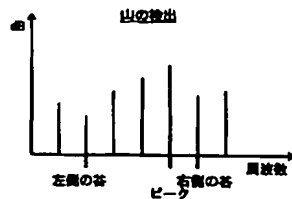
【図5】



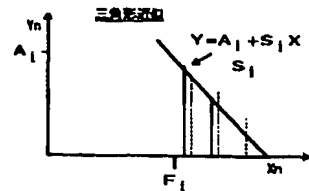
【図9】



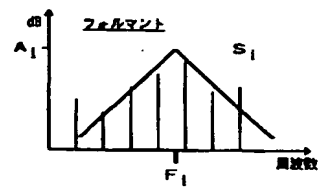
【図13】



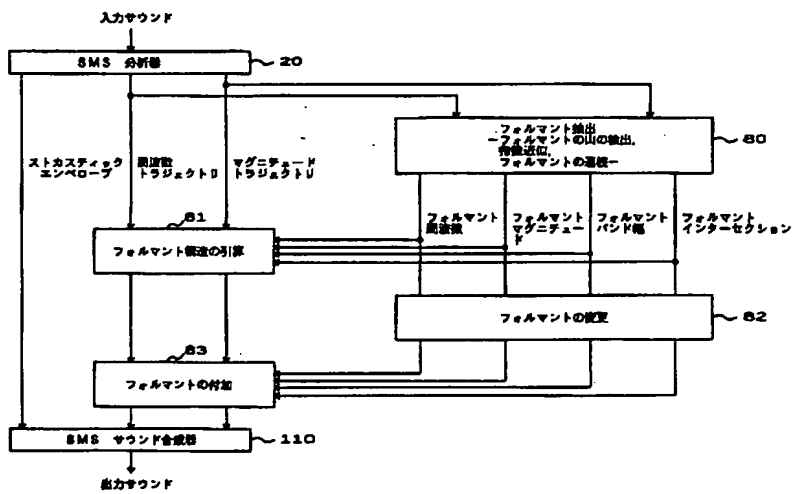
【図14】



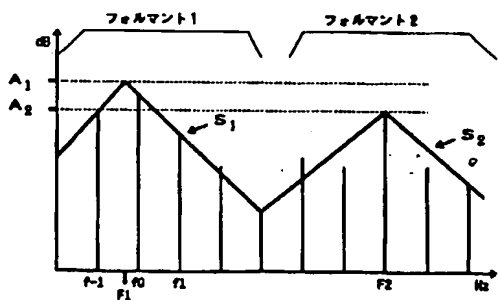
【図15】



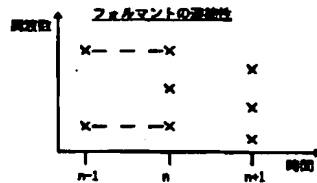
【図6】



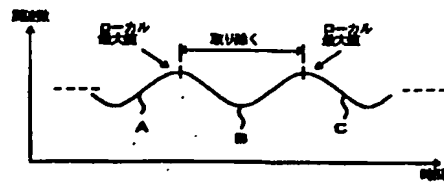
【図12】



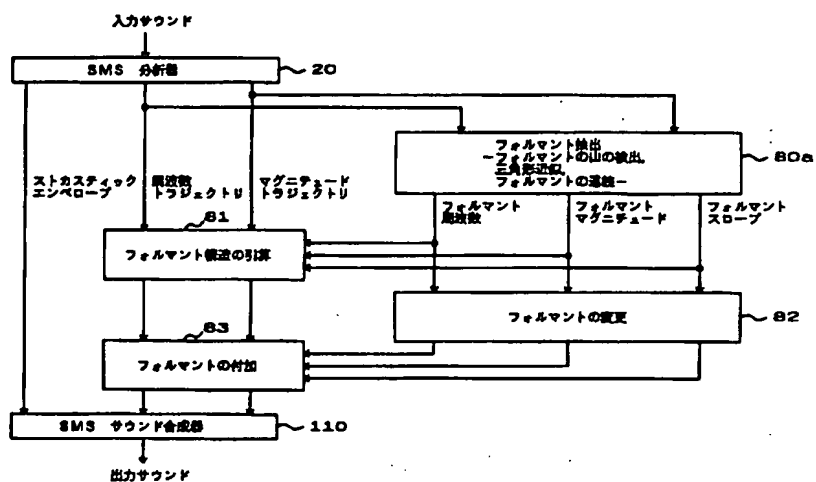
【図16】



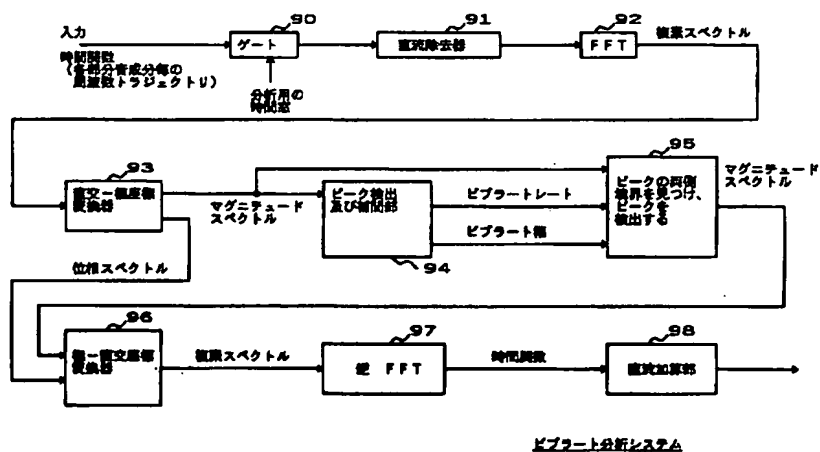
【図26】



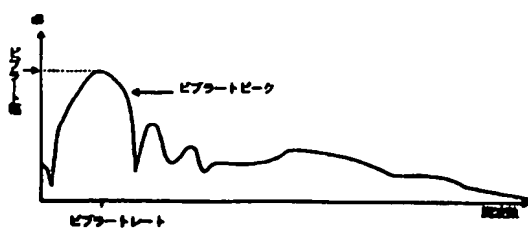
【图 1 1】



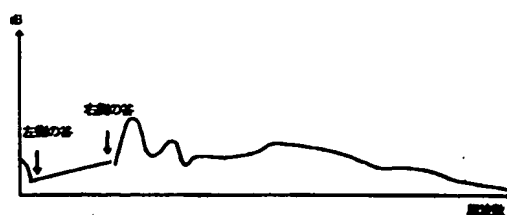
【图 1 7】



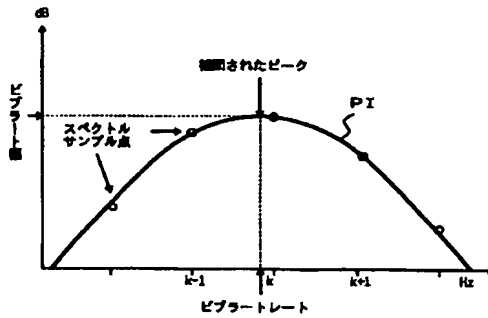
【图 18】



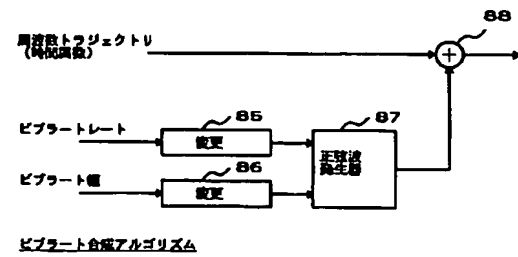
【图 19】



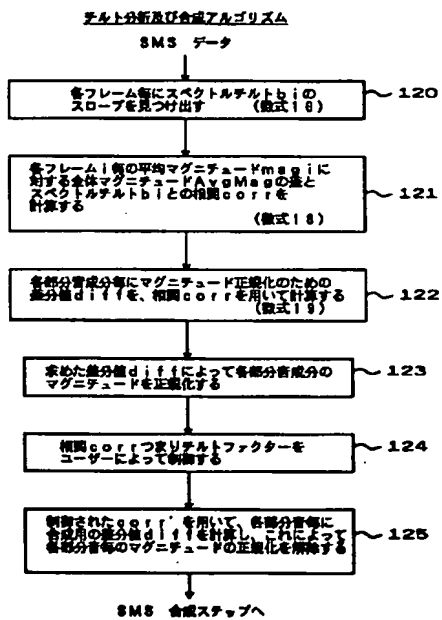
【図20】



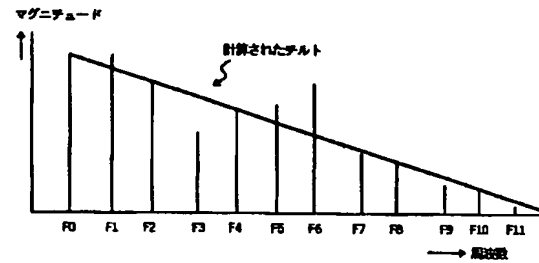
【図21】



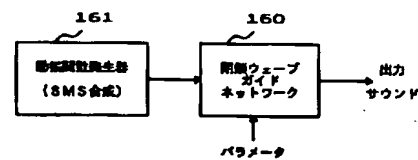
【図22】



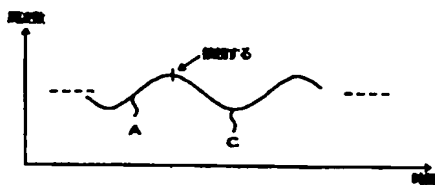
【図23】



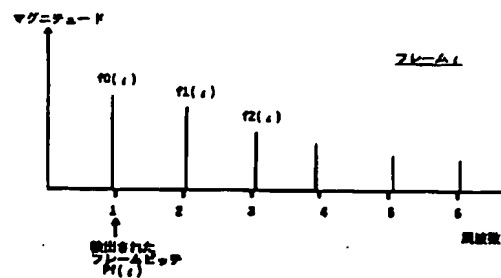
【図31】



【図27】



【図30】



```

graph TD
    130[130 最初のローカル最大値を見つけて出す] --> 131[131 値を見つけて出す]
    131 --> 132[132 値の確定]
    132 --> 133[133 をダント取り出す]
    133 --> 134[134 をダント取り出す]
    134 --> 135[135 処理地帯]
    135 --> 136[136 終了]
  
```

130 最初のローカル最大値を見つけて出す

131 値を見つけて出す

132 値の確定

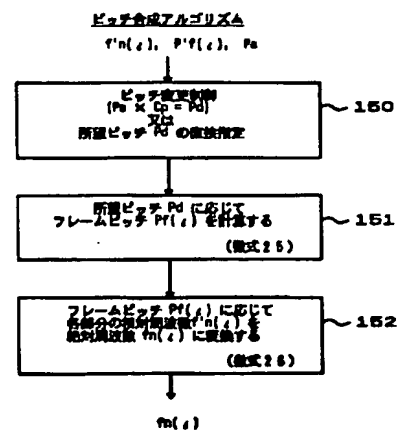
133 をダント取り出す

134 をダント取り出す

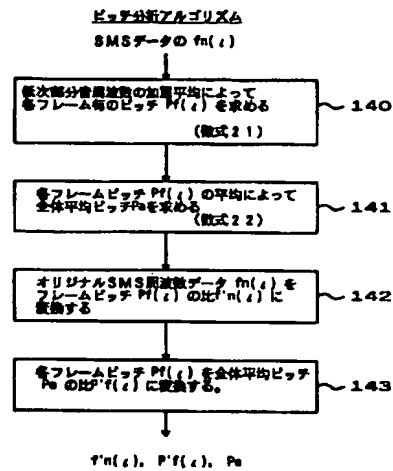
135 処理地帯

136 終了

【图 29】



【図28】



## フロントページの続き

(72)発明者    ロバート    グロス  
                  アメリカ合衆国, ノースカロライナ  
                  27615, ローリー, サウスフィールド  
                  ドライブ 8509  
 (72)発明者    アーリング    ウォルド  
                  アメリカ合衆国, カリフォルニア  
                  94530, エルサーリト, ルドウィグ    ア  
                  ペニュー    5618

(56)参考文献    特開    平1-209497 (JP, A)  
                  特開    昭60-97397 (JP, A)  
                  特開    昭63-61296 (JP, A)  
                  特開    平5-119782 (JP, A)  
                  特開    平5-127678 (JP, A)

(58)調査した分野(Int.Cl.<sup>6</sup>, DB名)  
                  G10H    7/08  
                  G10H    7/00



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**